

STIC-Biotech/ChemLib

88752

From: Davis, Minh-Tam
Sent: Tuesday, March 11, 2003 3:01 PM
To: STIC-Biotech/ChemLib
Subject: Search request for 09/847046

Please search in commercial database, PGPUB, and issued patent files:

1) SEQ ID NO:1 and 2.

Thank you.

MINH TAM DAVIS

ART UNIT 1642, ROOM 8A01, MB 8E12

305-2008

Oct 2000
party date

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

RECEIVED
MAR 11 2003
STIC/Biotech
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/13/03
Date Completed: 3/25/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 8/10/04
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (US)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2003, 04:22:07 ; Search time 959 Seconds
(without alignments)
10628.305 Million cell updates/sec

Title: US-09-847-046-1
Perfect score: 4526
Sequence: 1 gccgagtcgctgtgcgctgc.....cctcaaaaaaaaaaaaaa 4526

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_101002:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4525.6	100.0	4526	ABK92127	Prostate cancer-as
2	2811.6	62.1	2820	AAH15799	Human cDNA sequenc
3	2540.4	56.1	2905	ABV21967	Human prostate exp
4	2540.4	56.1	2905	ABV27799	Human prostate exp
5	776.2	17.1	784	AAH05323	Human cDNA clone (
6	349.8	7.7	360	ABV12212	Human prostate exp
7	348.8	7.7	374	ABV03043	Human prostate exp
8	348.8	7.7	397	ABV33357	Human prostate exp
9	348.8	7.7	397	ABV42280	Human prostate exp

10	341	7.5	355	21	AAK28836	Human secreted pro
11	307.6	6.8	416	23	ABV38274	Human prostate exp
12	304.4	6.7	416	23	ABV25366	Human prostate exp
13	290	6.4	547	22	AAK93519	cDNA encoding SKT
14	234.6	5.2	447	23	ABV08367	Human prostate exp
15	106	2.3	1423	22	AAK26179	Human cDNA encodin
16	97.8	2.2	6792	21	AAK94062	Arlyl hydrocarbon r
17	95.4	2.1	32986	22	AAK69758	Human immune/haema
18	95.4	2.1	32986	22	AAK84629	Human immune/haema
19	93.2	2.0	3420	22	AAH46280	Human proliferatin
20	92.6	2.0	34878	22	AAK6167	Human immune/haema
21	92.6	2.0	34878	22	AAK80088	Human immune/haema
22	91.4	2.0	17327	14	AAK04278	Serglycin - proteo
23	90.6	2.0	4971	22	ABA20298	Human nervous syst
24	90.2	2.0	90220	24	ABK83576	Human cDNA differe
25	86.8	1.9	811	22	AAK00199	Human reproductive
26	86.8	1.9	1456	22	ABV18636	Human nervous syst
27	86.8	1.9	31208	24	ABK52899	Genomic DNA encod1
28	86	1.9	143	22	AAK69245	Human cervical can
29	84.8	1.9	122186	22	AAK89560	Human histone deac
30	84.4	1.9	43069	21	AAK36335	Genomic sequence o
31	82.4	1.8	13712	24	ABL33531	Human immune syste
32	81.4	1.8	1083	22	AAH64963	Human secreted pro
33	79.8	1.8	546	22	ABA59802	Human foetal liver
34	79.8	1.8	546	22	AAK08075	Human brain expres
35	79.8	1.8	546	22	AAK33949	Human bone marrow
36	79.5	1.8	147309	24	ABK49450	Human transporter
37	79.2	1.7	149671	24	ABK84797	Human cDNA differe
38	79	1.7	5314	22	AAK64723	Human immune/haema
39	79	1.7	5314	22	AAK75611	Human immune/haema
40	79	1.7	5314	22	AAK91497	Human digestive sy
41	79	1.7	9947	24	ABK52825	Genomic DNA encod1
42	76.8	1.7	1808	21	AAK95521	Human secreted pro
43	76.2	1.7	34337	24	ABL55857	Human GABA transpo
44	76	1.7	13605	22	ABL37081	Human musculoskele
45	75.6	1.7	13712	24	ABL33530	Human immune syste

ALIGNMENTS

RESULT 1
ABK92127
ID ABK92127 standard; DNA: 4526 BP.
AC ABK92127;
XX
DT 15-AUG-2002 (first entry)
XX
DE Prostate cancer-associated DNA sequence #13.
XX
DE
XX
KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
KM gene therapy; gene; ds.
XX
OS Mammalia.
XX
PN WO200230268-A2.
PD
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US32045.
XX
PF 13-OCT-2000; 2000US-0687576.
PR 08-DEC-2000; 2000US-0733288.
PR 08-DEC-2000; 2000US-0733742.
PR 24-JAN-2001; 2001US-263957P.
PR 16-MAR-2001; 2001US-276791P.
PR 16-MAR-2001; 2001US-276791P.
PR 06-APR-2001; 2001US-281922P.
PR 24-APR-2001; 2001US-286214P.
PR 30-APR-2001; 2001US-0847046.
PR 04-MAY-2001; 2001US-288589P.
XX

OY	1621	TTAACTGAAAGGACACATCTACTATCTGCTCTTCTACTAGATTAACGTGTGACATGTGCTG	1680
Db	1621	TTAAACTGAAAGGACACATCTACTATCTGCTCTTCTACTAGATTAACGTGTGACATGTGCTG	1680
OY	1681	ATGTATGTACTAAGCAAAAGTTACTAGAAATTTCTCATCTAAAGATGTACAAAGAGCG	1740
Db	1681	ATGTATGTACTAAGCAAAAGTTACTAGAAATTTCTCATCTAAAGATGTACAAAGAGCG	1740
OY	1741	AGAACCCAGTATCTTATGCTGGATGTTAGGAACCAAGATCTCTAAATTTTATTCAGC	1800
Db	1741	AGAACCCAGTATCTTATGCTGGATGTTAGGAACCAAGATCTCTAAATTTTATTCAGC	1800
OY	1801	TCAACAGATTTTCATATTCAGGTAAATATACATCCAAACAGCAGAGATTTTAA	1860
Db	1801	TCAACAGATTTTCATATTCAGGTAAATATACATCCAAACAGCAGAGATTTTAA	1860
OY	1861	GTGGGGAATTTATTAAGACCTCATCTGATATCTATGCTGTGCAGATTTGGACATTTTA	1920
Db	1861	GTGGGGAATTTATTAAGACCTCATCTGATATCTATGCTGTGCAGATTTGGACATTTTA	1920
OY	1921	GTCCAACCATATGAAAAACAGAAAAAGATTTTATGTAACAGAGAACTACTTAAAGAT	1980
Db	1921	GTCCAACCATATGAAAAACAGAAAAAGATTTTATGTAACAGAGAACTACTTAAAGAT	1980
OY	1981	ATGTATATCATCTGGAATTTATTTCTGAAGAGATGTTTGTACTGTCAACCAATATGCTG	2040
Db	1981	ATGTATATCATCTGGAATTTATTTCTGAAGAGATGTTTGTACTGTCAACCAATATGCTG	2040
OY	2041	CAAGCTTCCAGCCCTGCTCTTGCACAGACACACAGAACGCAAAATATGAGACATCCAC	2100
Db	2041	CAAGCTTCCAGCCCTGCTCTTGCACAGACACACAGAACGCAAAATATGAGACATCCAC	2100
OY	2101	TAGCTAGCACACATGCAACAAGACATAGTTCAATATAAACAAGATCAGTCTGAAAAATG	2160
Db	2101	TAGCTAGCACACATGCAACAAGACATAGTTCAATATAAACAAGATCAGTCTGAAAAATG	2160
OY	2161	TTCCGGAAATCAGTGTGAAAAATCTTCCAGATTTTTCAGACTTCAGAAACCATATTGA	2220
Db	2161	TTCCGGAAATCAGTGTGAAAAATCTTCCAGATTTTTCAGACTTCAGAAACCATATTGA	2220
OY	2221	TTTTGTTCAGATGGGACATGTAATCTCCATATATTAACCAATATTTGACATCGTAA	2280
Db	2221	TTTTGTTCAGATGGGACATGTAATCTCCATATATTAACCAATATTTGACATCGTAA	2280
OY	2281	AGCAGAAATATCTGTGATTCATTACTTACCCATGCGTGTAAATTAAGAAATCTCCAGTGG	2340
Db	2281	AGCAGAAATATCTGTGATTCATTACTTACCCATGCGTGTAAATTAAGAAATCTCCAGTGG	2340
OY	2341	GGAGAGGAATCTTGCGGGCATATTTTGGATCTCTGTGCTCCCTCTCTGTGTTTGG	2400
Db	2341	GGAGAGGAATCTTGCGGGCATATTTTGGATCTCTGTGCTCCCTCTCTGTGTTTGG	2400
OY	2401	TGAATCTGCATTCAGGTGGCCAAATTTTTCGATTTCTTCAGACCAAGCCTTAATTGAG	2460
Db	2401	TGAATCTGCATTCAGGTGGCCAAATTTTTCGATTTCTTCAGACCAAGCCTTAATTGAG	2460
OY	2461	AAAACTTGATTTGTGGCTCAGAAATTTGAAGCAGGACTAGAAAAATCATATTCACAATTT	2520
Db	2461	AAAACTTGATTTGTGGCTCAGAAATTTGAAGCAGGACTAGAAAAATCATATTCACAATTT	2520
OY	2521	TACCTGCTCAAGATGGAACCTCTCTTCCAGCTTATGATTTTCTAAGTATGATAGATG	2580
Db	2521	TACCTGCTCAAGATGGAACCTCTCTTCCAGCTTATGATTTTCTAAGTATGATAGATG	2580
OY	2581	CCGCAACATCTCAACGTGGCAGCTAGGAAAGTTCCCAAGGTGTGAAAGAAACACATGTGC	2640
Db	2581	CCGCAACATCTCAACGTGGCAGCTAGGAAAGTTCCCAAGGTGTGAAAGAAACACATGTGC	2640
OY	2641	AGGAGAAATGATTAAGGAACACATGGAAGATTAATGGCAGCTCGAAAGAAACCCGATTGAA	2700
Db	2641	AGGAGAAATGATTAAGGAACACATGGAAGATTAATGGCAGCTCGAAAGAAACCCGATTGAA	2700
OY	2701	CTTGACAAATTAAGCATGTGGAATAGGAATTAATTTGTTTAAAGAACAGAAAAATCATTTTA	2760

Db	2701	CTCTGAGAAATAAACCATTTGGAAATGAAGTATTTGGTTTAAAGACGACGAAAAATCATTTTAA	2760
Qy	2761	GACGGATAAAGAGATTTAGATGCTCAAAAGTGAACATAATTTATAGGGCTGTGGTTTCCA	2820
Db	2761	GACGGATAAAGAGATTTAGATGCTCAAAAGTGAACATAATTTATAGGGCTGTGGTTTCCA	2820
Qy	2821	AAATTTTGGCATGTATGATCTTAATTTATTTCCTTAAAGAAATATATTTAAATCATTTTC	2880
Db	2821	AAATTTTGGCATGTATGATCTTAATTTATTTCCTTAAAGAAATATATTTAAATCATTTTC	2880
Qy	2881	AAGTTTGACAGCATGTGCGCATCCCAATGATTTATATATAGTCACATRTTTTATTTAA	2940
Db	2881	AAGTTTGACAGCATGTGCGCATCCCAATGATTTATATATAGTCACATRTTTTATTTAA	2940
Qy	2941	ATTTTCTGTAACTACATATTTAAACAAAGTAAAGTGAAGGCGCAAAATTAATTTGATTT	3000
Db	2941	ATTTTCTGTAACTACATATTTAAACAAAGTAAAGTGAAGGCGCAAAATTAATTTGATTT	3000
Qy	3001	ACTTTTCACCCAGTATGATATACCCAAATATGCGAAATATATGAATTTATTAATGATATTT	3060
Db	3001	ACTTTTCACCCAGTATGATATACCCAAATATGCGAAATATATGAATTTATTAATGATATTT	3060
Qy	3061	TGACATCCTTTTGTACCAAGTCTCTTAATGACATGATTTTATATCTTATCTGATTT	3120
Db	3061	TGACATCCTTTTGTACCAAGTCTCTTAATGACATGATTTTATATCTTATCTGATTT	3120
Qy	3121	CTTACTTCCGAGTGGCCATATTTCAAGTCTTCAATGCCACATGTGGCCGTGACTACGT	3180
Db	3121	CTTACTTCCGAGTGGCCATATTTCAAGTCTTCAATGCCACATGTGGCCGTGACTACGT	3180
Qy	3181	ATTGACAGTTCAGTACTAGACAAAACTGACATAATTAATTAATCTTACGACATGATTT	3240
Db	3181	ATTGACAGTTCAGTACTAGACAAAACTGACATAATTAATTAATCTTACGACATGATTT	3240
Qy	3241	CTATTTGGATTTAAATTTAACTCTAATACAGTTAACTCCACAGTGCATTCATGACGTG	3300
Db	3241	CTATTTGGATTTAAATTTAACTCTAATACAGTTAACTCCACAGTGCATTCATGACGTG	3300
Qy	3301	ACAGTTAATATTTGTTTTATTTGAGAGCTGATTTAAATACAGCGTTTGTCAACCTCAGG	3360
Db	3301	ACAGTTAATATTTGTTTTATTTGAGAGCTGATTTAAATACAGCGTTTGTCAACCTCAGG	3360
Qy	3361	GATATTTAGCAATTTGTGCGGAGACATTTTGTGATCATACATAGAGGCAAGTATTTGACATTT	3420
Db	3361	GATATTTAGCAATTTGTGCGGAGACATTTTGTGATCATACATAGAGGCAAGTATTTGACATTT	3420
Qy	3421	TAGTGAGTAGAGGCCATGATCCTGCTAAATTAACCTGCATTTGAGACGCGCCACACAA	3480
Db	3421	TAGTGAGTAGAGGCCATGATCCTGCTAAATTAACCTGCATTTGAGACGCGCCACACAA	3480
Qy	3481	AGAATTAATCCCGCCGAAATGTATGTGTCGCGCAAGGCTAGTATACCTGTGTAAAGTA	3540
Db	3481	AGAATTAATCCCGCCGAAATGTATGTGTCGCGCAAGGCTAGTATACCTGTGTAAAGTA	3540
Qy	3541	ACCTGTGCGACAGTACGATTTCCAGAAATTTCTGATCTGCTACGATCATGTTTGAATA	3600
Db	3541	ACCTGTGCGACAGTACGATTTCCAGAAATTTCTGATCTGCTACGATCATGTTTGAATA	3600
Qy	3601	AAATTTGGCTAATTAAGATATGTATTAGATGATCTTATTCCTGATTTATACCTGATCA	3660
Db	3601	AAATTTGGCTAATTAAGATATGTATTAGATGATCTTATTCCTGATTTATACCTGATCA	3660
Qy	3661	CTTGATCTTTTCTAATTTTTCAGAAAGTATGAGGATTAACCTACAAAGAGATCTCAGAT	3720
Db	3661	CTTGATCTTTTCTAATTTTTCAGAAAGTATGAGGATTAACCTACAAAGAGATCTCAGAT	3720
Qy	3721	GATATTTATTTATTTAAGTGAAGTCTTAAACCTGCTTATTTCTCAAGTTATATGAGCTA	3780
Db	3721	GATATTTATTTATTTAAGTGAAGTCTTAAACCTGCTTATTTCTCAAGTTATATGAGCTA	3780
Qy	3781	AAATTCAGATTTGAACAGGATTCAGCATTCCTGCAATCTCTCATGAAAGAGAGCTCC	3840
Db	3781	AAATTCAGATTTGAACAGGATTCAGCATTCCTGCAATCTCTCATGAAAGAGAGCTCC	3840

```
Db 3781 AATTTGAGATTGAAAGGAGTTCAGACATCTGCCATCTCCTATGGAAGAAGGCTCC 3840
Qy 3841 TCATCTGAAGCGTCTCTGGAATCTACCTTGAAGCTTACAGAAATCAGTATCTCC 3900
Db 3841 TCATCTGAAGCGTCTCTGGAATCTACCTTGAAGCTTACAGAAATCAGTATCTCC 3900
Qy 3901 TGAGCCACACGCGCTCTCTGAGAGGAGGAAAGATTAGCCAAAGAGTTAATTTTCAT 3960
Db 3901 TGAGCCACACGCGCTCTCTGAGAGGAGGAAAGATTAGCCAAAGAGTTAATTTTCAT 3960
Qy 3961 TCAAAATCACTTACCTGTTAGACGATCTGTTGACAGTTGTTGTTTCATTTTTCAT 4020
Db 3961 TCAAAATCACTTACCTGTTAGACGATCTGTTGACAGTTGTTGTTTCATTTTTCAT 4020
Qy 4021 CTGTGCAATTTTGTGACATTTGTTGAAATATTTCTATTTGGTCTCTACTGATTTTTC 4080
Db 4021 CTGTGCAATTTTGTGACATTTGTTGAAATATTTCTATTTGGTCTCTACTGATTTTTC 4080
Qy 4081 TTTTAAATCTACTGATATCTTGTCTTAAATTTTCTTACATATGCTTGGCTGAT 4140
Db 4081 TTTTAAATCTACTGATATCTTGTCTTAAATTTTCTTACATATGCTTGGCTGAT 4140
Qy 4141 ACAAGTATTTTATACGTAATTAAGAAATCTAACAGTAAACTCAGTAAGTGCAT 4200
Db 4141 ACAAGTATTTTATACGTAATTAAGAAATCTAACAGTAAACTCAGTAAGTGCAT 4200
Qy 4201 MTATTTCTTATACATAGACCCGTTGCTACTCTCAGACACCTCTCTCTCAATTTTTC 4260
Db 4201 MTATTTCTTATACATAGACCCGTTGCTACTCTCAGACACCTCTCTCTCAATTTTTC 4260
Qy 4261 CTGTGACATGATGCTGATTAATACATTTTCATTTGCTTTTATTTCTAATATGGA 4320
Db 4261 CTGTGACATGATGCTGATTAATACATTTTCATTTGCTTTTATTTCTAATATGGA 4320
Qy 4321 CAATGAGAGTGAATCTAATATATAGTTAGTAAATCAATCATTAGGCTAATTTTA 4380
Db 4321 CAATGAGAGTGAATCTAATATATAGTTAGTAAATCAATCATTAGGCTAATTTTA 4380
Qy 4381 GAAATGCTAATTAAGTACGACACATGAACATGAATGCTTACTGATTTGACCTTT 4440
Db 4381 GAAATGCTAATTAAGTACGACACATGAACATGAATGCTTACTGATTTGACCTTT 4440
Qy 4441 GTGAGCAATTTTGCAGTCAATTTGTTGCTCAATTAATTAATGAATGCTGCGTTTC 4500
Db 4441 GTGAGCAATTTTGCAGTCAATTTGTTGCTCAATTAATTAATGAATGCTGCGTTTC 4500
Qy 4501 AGAATACCTTCAAAAAA 4526
Db 4501 AGAATACCTTCAAAAAA 4526

RESULT 2
AAH15799
ID AAH15799 standard; cDNA; 2820 BP.
XX
AC AAH15799;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14261.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
```

```
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 8; SEQ ID 14261; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides and the combination
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB9246 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2820 BP; 911 A; 496 C; 504 G; 909 T; 0 other;
XX
Query Match 62.1%; Score 2811.6; DB 22; Length 2820;
Best Local Similarly 99.8%; Pred. No. 0;
Matches 2814; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Qy 1686 TGTTACTAAGCAAAATGTTACTGTAATTTCTATCATTAAGATGTACAGAAAGCGAGAAC 1745
Db 1 TGTTACTAAGCAAAATGTTACTGTAATTTCTATCATTAAGATGTACAGAAAGCGAGAAC 60
Qy 1746 CCAGTATCTTATGCTGGAATGTTAGAACCAAGATCTCCTAAATTTATCCAGCTCAAC 1805
Db 61 CCAGTATCTTATGCTGGAATGTTAGAACCAAGATCTCCTAAATTTATCCAGCTCAAC 120
Qy 1806 AGGATTTCAATATCAGTGAATTAACATGATCCAAAGACAGAAAGATTTTAATGAG 1865
Db 121 AGGATTTCAATATCAGTGAATTAACATGATCCAAAGACAGAAAGATTTTAATGAG 180
Qy 1866 GAATTTATTAAGACCCATCTGTAATTCAGTGCAGTATTTGGAGCATTTATTCACA 1925
Db 181 GAATTTATTAAGACCCATCTGTAATTCAGTGCAGTATTTGGAGCATTTATTCACA 240
Qy 1926 ACCATGAAAACAGCAAAAGAGATTTTAGTGAAGACGAAAGATCTTAAAGAGATGTT 1985
Db 241 ACCATGAAAACAGCAAAAGAGATTTTAGTGAAGACGAAAGATCTTAAAGAGATGTT 300
Qy 1986 ATCACTGGAATTTATCTGAAGAAGATGTTTCTACTGTCACCAAAATATGCTGCAAGT 2045
Db 301 ATCACTGGAATTTATCTGAAGAAGATGTTTCTACTGTCACCAAAATATGCTGCAAGT 360
Qy 2046 CTTCACAGCCCTGCTGCTGCGACACACGACGAAAGCAAAATAGAGCATCCACATGCT 2105
Db CTTCACAGCCCTGCTGCTGCGACACACGACGAAAGCAAAATAGAGCATCCACATGCT 2105
```

Db	361	CTTCACAGCCCTGCTGCTTGGCCACACACACAGAAGCCAAAATTAGAGAGATCCCACTAGCT	420
Qy	2106	AGCACACATGCACANAGACATAGTTCTCAAAATPAATPAACGATGCTACTGGAATGTTC	2165
Db	421	AGCACACATGCACANAGACATAGTTCTCAAAATPAATPAACGATGCTACTGGAATGTTC	480
Qy	2166	GAATACCTGTGAAAATCTGCCACTATTTTCAGACTTCAGAAACCATATATGATTTTG	2225
Db	481	GAATACCTGTGAAAATCTGCCACTATTTTCAGACTTCAGAAACCATATATGATTTTG	540
Qy	2226	TTTCAGTATGGCACTGTAAATTCCTCAATATTAATAAAGCAATTTGACACTGTTAAACG	2285
Db	541	TTTCAGTATGGCACTGTAAATTCCTCAAGTATTAATAAAGCAATTTGACACTGTTAAACG	600
Qy	2286	AAATACCTGGATCATATTACTCATCTGTTAAATCTPAACAAATACCCAGTGGGAGA	2345
Db	601	AAATACCTGGATCATATTACTCATCTGTTAAATCTPAACAAATACCCAGTGGGAGA	660
Qy	2346	GGAATCTTGGGGGCATATTTTGATTCCTTGCCCTCCCTTCCTCTTGTTGGTGGAT	2405
Db	661	GGAATCTTGGGGGCATATTTTGATTCCTTGCCCTCCCTTCCTCTTGTTGGTGGAT	720
Qy	2406	CTGATTCAGTGGGCCAAGATTTGATTTCCCTCAGACACAGCTATATTTGAGAAGAAC	2465
Db	721	CTGATTCAGTGGGCCAAGATTTGATTTCCCTCAGACACAGCTATATTTGAGAAGAAC	780
Qy	2466	CTTGTATTGTGGCGTGAAGAAATTAGAGAGAGACTAGAAAAATCATATCACATTTTACCT	2525
Db	781	CTTGTATTGTGGCGTGAAGAAATTAGAGAGAGACTAGAAAAATCATATTCATTTTACCT	840
Qy	2526	GCTCAGAAATGGAACCTCCTCTTCAGCTTATGATTTTCTAAGTATGATAGTCCGCA	2585
Db	841	GCTCAGAAATGGAACCTCCTCTTCAGCTTATGATTTTCTAAGTATGATAGTCCGCA	900
Qy	2586	ACATCTCAACGTGGCACTAGSAAAGTTCCCAAGTATGSAAGAAACAGATGTGCAGAG	2645
Db	901	ACATCTCAACGTGGCACTAGSAAAGTTCCCAAGTATGSAAGAAACAGATGTGCAGAG	960
Qy	2646	AATATATTAAGAACACACATGAGATTAATCGGACGTGAGAAAAGAACGATTTGAAACTGTG	2705
Db	961	AATATATTAAGAACACACATGAGATTAATCGGACGTGAGAAAAGAACGATTTGAAACTGTG	1020
Qy	2706	AGAAATTAAGCATTTGGAATAGAACTAATTTGTTTAAAGAACGAAAATTCATTTAGACGT	2765
Db	1021	AGAAATTAAGCATTTGGAATAGAACTAATTTGTTTAAAGAACGAAAATTCATTTAGACGT	1080
Qy	2766	GATTAAGATTTAGAGATGCTCAAAAAGTCAATTTTATAGGGCTGTGTTCCAAAATT	2825
Db	1081	GATTAAGATTTAGAGATGCTCAAAAAGTCAATTTTATAGGGCTGTGTTCCAAAATT	1140
Qy	2826	TTTTTGGCATGATAGACTTAATTTATTTTCCCTTAAGAAATPAATPAATCAATTCAGATT	2885
Db	1141	TTTTTGGCATGATAGACTTAATTTATTTTCCCTTAAGAAATPAATPAATCAATTCAGATT	1200
Qy	2886	TGCAGACTAGTGCATTCACATAGAAATTAATPAATPAAGTCCATATTTTATTTAAATTTT	2945
Db	1201	TGCAGACTAGTGCATTCACATAGAAATTAATPAATPAAGTCCATATTTTATTTAAATTTT	1260
Qy	2946	CTATATCTACATTTAAACAAAGTAAAGAGACAGGGCAAAATATTTTGATATTACTTT	3005
Db	1261	CTATATCTACATTTAAACAAAGTAAAGAGACAGGGCAAAATATTTTGATATTACTTT	1320
Qy	3006	TCACCCAGTGTATACCCAAATAGGGAATATAGAAATTAATTAAGATATTTTACAT	3065
Db	1321	TCACCCAGTGTATACCCAAATAGGGAATATAGAAATTAATTAAGATATTTTACAT	1380
Qy	3066	CCTTTTTTGTACCAAGTCTCTAAATGACGATACATATTTTATATCTACGATTTCTTAC	3125
Db	1381	CCTTTTTTGTACCAAGTCTCTAAATGACGATACATATTTTATATCTACGATTTCTTAC	1440
Qy	3126	TTCCGAGTAGGCATATTTCAAGTGTTCATTTGCCACATGTGGCTGTGACTACTGATTTGG	3185
Db	1441	TTCCGAGTAGGCATATTTCAAGTGTTCCTTGGCACATGTGGCTGTGACTACTGATTTGG	1500

QY	3186	ACAATTCAGTACTAGACAAAAAAGTACGATATTAATTAAGTTCTAGGCCATGATTTCTATT	3245
Db	1501	ACAGTTTACGACTAGACAAAAAAGTACGATATTAATTAAGTTCTAGGCCATGATTTCTATT	1560
QY	3246	TGGATTAAATTAACCTTAATACAGTTTAAGTCTCCAGAGTATTCATGACAGCTGCACAGT	3305
Db	1561	TGGATTAAATTAACCTTAATACAGTTTAAGTCTCCAGAGTATTCATGACAGCTGCACAGT	1620
QY	3306	TATATTTGTTTTTTTGGAGCATGATATTAATAATCAGCTTTGTCACCTCAGGAGATAT	3365
Db	1621	TATATTTGTTTTTTTGGAGCATGATATTAATAATCAGCTTTGTCACCTCAGGAGATAT	1680
QY	3366	TTAGCAATGTCGGAGACATTTTTTGATGTCATGACTAGACGGCAGTTATTTGACATTTAGTG	3425
Db	1681	TTAGCAATGTCGGAGACATTTTTTGATGTCATGACTAGACGGCAGTTATTTGACATTTAGTG	1740
QY	3426	AGTAGAGCCATGGATCTCTCTAATAATTAACCTGCATTTGGACAGCCGCCACAACAAGAT	3485
Db	1741	AGTAGAGCCCAAGGATCTCTCTAATAATTAACCTGCATTTGGACAGCCGCCACAACAAGAT	1800
QY	3486	TATCTCCCGGAAATGGTAGTCTGGCCAAAGGCGAGTAACTGTGTTAAAAGTAACCTG	3545
Db	1801	TATCTCCCGGAAATGGTAGTCTGGCCAAAGGCGAGTAACTGTGTTAAAAGTAACCTG	1860
QY	3546	TGGCAGACTAGTTTCCAGAAATTTCCGTGTTGCTGCACAGTATCATGTTTGAAAAAATTT	3605
Db	1861	TGGCAGACTAGTTTCCAGAAATTTCCGTGTTGCTGCACAGTATCATGTTTGAAAAAATTT	1920
QY	3606	TGCGTATTAAGATATGATATTTAGATGCTCTATCTGATTTATTAACCTGCATACACTTGA	3665
Db	1921	TGCGTATTAAGATATGATATTTAGATGCTCTATCTGATTTATTAACCTGCATACACTTGA	1980
QY	3666	TCTTTTCTAATATTTTCAGAAAGTGATGGGATAAACCCTTAGAAGAGAGACTCGAATGATAT	3725
Db	1981	TCTTTTCTAATATTTTCAGAAAGTGATGGGATAAACCCTTAGAAGAGAGACTCGAATGATAT	2040
QY	3726	TTATATTTTAAGTGAGCTTAAACACCTCCCTTATTTCTACAAAGTTATATGAGCTAAATTT	3785
Db	2041	TTATATTTTAAGTGAGCTTAAACACCTCCCTTATTTCTACAAAGTTATATGAGCTAAATTT	2100
QY	3786	CAGATTGAACAGGGATTCAGACATTTGCGCATCTCCTCATGAAAGAGAGGCTCCCTCATC	3845
Db	2101	CAGATTGAACAGGGATTCAGACATTTGCGCATCTCCTCATGAAAGAGAGGCTCCCTCATC	2160
QY	3846	TGAAGCCTCTCTGAAATCTACCTCTGCAAGCTTCAGACAAATCACTTATCTCCCTGAGC	3905
Db	2161	TGAAGCCTCTCTGAAATCTACCTCTGCAAGCTTCAGACAAATCACTTATCTCCCTGAGC	2220
QY	3906	CACACGGCCTCATCTCTGAGGAGGAGAAAGATTAAGCCAAAGATTAATTTTCATTCGAA	3965
Db	2221	CACACGGCCTCATCTCTGAGGAGGAGAAAGATTAAGCCAAAGATTAATTTTCATTCGAA	2280
QY	3966	ATCATTAGCTGTTAAGACATGATCTGTTTGTAGCAGTTGTTTGTCTCATATTTTGTCTGTG	4025
Db	2281	ATCATTAGCTGTTAAGACATGATCTGTTTGTAGCAGTTGTTTGTCTCATATTTTGTCTGTG	2340
QY	4026	CATTTTGGAGCAATTTGTTGAATAATTCATTTGAGTGTCTCTACTGATATTTTCTTTTTT	4085
Db	2341	CATTTTGGAGCAATTTGTTGAATAATTCATTTGAGTGTCTCTACTGATATTTTCTTTTTT	2400
QY	4086	AATATCTACTGATATCTTCTTTTAAATTTCTTCACATATATGTTTGGCTGCATACAC	4145
Db	2401	AATATCTACTGATATCTTCTTTTAAATTTCTTCACATATATGTTTGGCTGCATACAC	2460
QY	4146	TGATTTTATTAACGAAATTTAAGAACTTAACAGCTTAAAACTCAGTAAGTGCATMTATT	4205
Db	2461	TGATTTTATTAACGAAATTTAAGAACTTAACAGCTTAAAACTCAGTAAGTGCATMTATT	2520
QY	4206	TCCTTTAATAATAGACCCGTTGCTACTCTCAGCACCCCTCTCTCAATTTTTTTTCTGTGA	4265
Db	2521	TCCTTTAATAATAGACCCGTTGCTACTCTCAGCACCCCTCTCTCAATTTTTTTTCTGTGA	2580

QY 4266 GCATGTCATGCTGATTAACTCATTTTCCTTTATTTCTAATATGGAACAATG 4325
|||||
DB 2581 GCATGTCATGCTGATTAACTCATTTTCCTTTATTTCTAATATGGAACAATG 2640
QY 4326 AGAGTGAACCTTAATATAGTGTGTAGTATTAACAACCTCTTAGCCCAATATATAGAAA 4385
|||||
DB 2641 AGAGTGAACCTTAATATAGTGTGTAGTATTAACAACCTCTTAGCCCAATATATAGAAA 2700
QY 4386 TGCATTAATAGTACGACGACATGAATGAATTCCTAGTATGTCATCTTGTGACG 4445
|||||
DB 2701 TGCATTAATAGTACGACGACATGAATGAATTCCTAGTATGTCATCTTGTGACG 2760
QY 4446 CAATTTTGAACGTCATTAATGTTTGTCAATATTTTAAATAAAGTGTCTGGGTTTCAAGAT 4505
|||||
DB 2761 CAATTTTGAACGTCATTAATGTTTGTCAATATTTTAAATAAAGTGTCTGGGTTTCAAGAT 2820

RESULT 3
ABV21967
ID ABV21967 standard; cDNA; 2905 BP.
XX
AC ABV21967;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 21958.
XX
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-25281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer; useful
PT for detecting presence of prostate cancer; stage of prostate cancer -
XX
XX Claim 1; Page 3747; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient.
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX

SQ Sequence 2905 BP; 921 A; 509 C; 554 G; 919 T; 2 other:
Query Match 56.1%; Score 2540.4; DB 23; Length 2905;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2627; Conservative 1; Mismatches 62; Indels 10; Gaps 4;
QY 1664 CTGTGAGATTTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAATTTCTATCATATA 1723
| | | | |
DB 18 CCGTCGAGATTTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAATTTCTATCATATA 77
QY 1724 GATGTACAAAGAGGCGAGAACCCAGATCTTATGCTGGAAATTTGGAAACCGAAGATCT 1783
| | | | |
DB 78 GATGTACAAAGAGGCGAGAACCCAGATCTTATGCTGGAAATTTGGAAACCGAAGATCT 137
QY 1784 CCTAAATTTATTCAGTCCACAGATTTTCATATCCAGTAATATATACATCATGATCCAGAGA 1843
| | | | |
DB 138 CCTAAATTTATTCAGTCCACAGATTTTCATATCCAGTAATATATACATCATGATCCAGAGA 197
QY 1844 AGCAGAAAGATTTTAAAGTGGGCAATTTATTAAGACCTCATCTTGTATTTCTAGTGTCT 1903
| | | | |
DB 198 AGCAGAAAGATTTTAAAGTGGGCAATTTATTAAGACCTCATCTTGTATTTCTAGTGTCT 257
QY 1904 AGTATTGGGACTATTTAGTCCACATGAACAGCAAAAGAAATTTTGTAGAGCAGG 1963
| | | | |
DB 258 AGTATTGGGACTATTTAGTCCACATGAACAGCAAAAGAAATTTTGTAGAGCAGG 317
QY 1964 AACTACTTAAAGAGATATGTATACCTGGAATTTATTTGAAAGAAATGTTTGTCTACT 2023
| | | | |
DB 318 AACTACTTAAAGAGATATGTATACCTGGAATTTATTTGAAAGAAATGTTTGTCTACT 377
QY 2024 GTCAACCAATATATGCTGCAAGTCTTCCAGCCCTGCTGCTTCCAGACACACAAAGGCA 2083
| | | | |
DB 378 GTCAACCAATATATGCTGCAAGTCTTCCAGCCCTGCTGCTTCCAGACACACAAAGGCA 437
QY 2084 ATATGAGAGCATCCCATCTAGTACACATGACAAAGACATATGTTCAATTAATTAACAGA 2143
| | | | |
DB 438 ATATGAGAGCATCCCATCTAGTACACATGACAAAGACATATGTTCAATTAATTAACAGA 497
QY 2144 TGCATCTAGTGAATGTTTCCGGAATCACTGTGGAATCTTCCAGTATTTTCAGACT 2203
| | | | |
DB 498 TGCATCTAGTGAATGTTTCCGGAATCACTGTGGAATCTTCCAGTATTTTCAGACT 557
QY 2204 TCAGAAACCTATTTGATTTTGTTCAGTGTATGTCACCTGTAAATCTTCAATTAATAAAGC 2263
| | | | |
DB 558 TCAGAAACCTATTTGATTTTGTTCAGTGTATGTCACCTGTAAATCTTCAATTAATAAAGC 617
QY 2264 AATATTGACATGCTGAAGCAGAAATCACTGATTCATTTACTCCATGCTGTTAATCT 2323
| | | | |
DB 618 AATATTGACATGCTGAAGCAGAAATCACTGATTCATTTACTCCATGCTGTTAATCT 677
QY 2324 AAAGATTAATCTCAGTGGGAGAGGAATCTTCCGGGCAATTTTGAATCTCTGCTCCCT 2383
| | | | |
DB 678 AAAGATTAATCTCAGTGGGAGAGGAATCTTGAAGGCAATTTTGAATCTCTGCTCCCT 737
QY 2384 TCCTCTTCTTGTGTTGGTGATCTGCATTCAGTGTGCGCAAGTATTTGCAATTTCTTCAGA 2443
| | | | |
DB 738 TCCTCTTCTTGTGTTGGTGATCTGCATTCAGTGTGCGCAAGTATTTGCAATTTCTTCAGA 797
QY 2444 CCAGGCTATTAATGAAGAAACCTTGATTTGAGGCTGAAGAAATTTGAAGCAGACTAGA 2503
| | | | |
DB 798 CCAGGCTATTAATGAAGAAACCTTGATTTGAGGCTGAAGAAATTTGAAGCAGACTAGA 857
QY 2504 AATCATATACCAATTTTACCTGCTCAGAGATGAAGAAACCTCCTCCAGCTATATAT 2563
| | | | |
DB 858 AATCATATACCAATTTTACCTGCTCAGAGATGAAGAAACCTCCTCCAGCTATATAT 917
QY 2564 TCTAATATGATATGATGTCGCAACATCTCAACGTCGACATAGAAAGTTCCAGGTAT 2623
| | | | |
DB 918 TCTAATATGATATGATGTCGCAACATCTCAACGTCGACATAGAAAGTTCCAGGTAT 977
QY 2624 GAAAGAAACAGATGTCAGAGAGATATGAAGAAACAAATGAAGATTAATTCGACATGCG 2683
| | | | |
DB 978 GAAAGAAACAGATGTCAGAGAGATATGAAGAAACAAATGAAGATTAATTCGACATGCG 1037

OY 2684 AAAAGAACCGATTGAAACTCTGAGATTAAGCATTTGGAATAGAGTAATTTGGTTAAAGA 2743
 DB 1038 AAAAGAACCGATTGAAACTCTGAGATTAAGCATTTGGAATAGAGTAATTTGGTTAAAGA 1097
 OY 2744 AGCAGAAAAATCATTTAGACGTGATTAAGAGTTAGAGTCTCMAAAGTGAAGTAATTTTA 2803
 DB 1098 AGCAGAAAAATCATTTAGACGTGATTAAGAGTTAGAGTCTCMAAAGTGAAGTAATTTTA 1157
 OY 2804 TAGGGCTGTGGTTCCAAAATTTTGGCATGATAGACTTAATTTATTTCTTAAAGAA 2863
 DB 1158 TAGGGCTGTGGTTCCAAAATTTTGGCATGATAGACTTAATTTATTTCTTAAAGAA 1217
 OY 2864 TAAATTAATCAATTTGCAAGTTGGAGACTAGTCCATCCATAGATTAATTAATTAAGT 2923
 DB 1218 TAAATTAATCAATTTGCAAGTTGGAGACTAGTCCATCCATAGATTAATTAATTAAGT 1277
 OY 2924 CACATATTTTATTTAAATTTTCTAGTAACCTACATTAAACAAAGTAAGAGCAGGCG 2983
 DB 1278 CACATATTTTATTTAAATTTTCTAGTAACCTACATTAAACAAAGTAAGAGCAGGCG 1337
 OY 2984 AAAAATAATTTGATATTTACTTTTCCAGTGTATACCCAAATATGCGAAATATAGAA 3043
 DB 1338 AAAAATAATTTGATATTTACTTTTCCAGTGTATACCCAAATATGCGAAATATAGAA 1397
 OY 3044 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTCTAAATGCAATGATAT 3103
 DB 1398 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTCTAAATGCAATGATAT 1457
 OY 3104 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTCTAAATGCAATGATAT 3163
 DB 1458 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTCTAAATGCAATGATAT 1517
 OY 3164 TGGCGTGTACGACGTGATTTGAGACAGTGTACAGTAAGAAAAATGATGATATTAAGT 3223
 DB 1518 TGGCGTGTACGACGTGATTTGAGACAGTGTACAGTAAGAAAAATGATGATATTAAGT 1577
 OY 3224 AGTTTACGATGATTTTCTATTTGATTAATAATTAACCTATGACAGTAACTGACA 3283
 DB 1578 AGTTTACGATGATTTTCTATTTGATTAATAATTAACCTATGACAGTAACTGACA 1637
 OY 3284 GTGCATTCATGACGTGACAGTATTAATTTGTTTATTTGAGTCAATGATTAATAATTCACG 3343
 DB 1638 GTGCATTCATGACGTGACAGTATTAATTTGTTTATTTGAGTCAATGATTAATAATTCACG 1697
 OY 3344 GTTGTCAACCTCAGGGGATTTTGTAGCAATGTCCGGAGACATTTTGTATGTCATGACATA 3403
 DB 1698 GTTGTCAACCTCAGGGGATTTTGTAGCAATGTCCGGAGACATTTTGTATGTCATGACATA 1757
 OY 3404 GGGCAGTTATTTGACATTTAGTGTAGAGGCCATGATCCTGCTAAATTAACCTGCAATGG 3463
 DB 1758 GGGCAGTTATTTGACATTTAGTGTAGAGGCCATGATCCTGCTAAATTAACCTGCAATGG 1817
 OY 3464 ACAGCGCCCCACACAAAGAAATTAATCTGCCCCGAAATGAGTGTGCTGCCAAGCGTAGTA 3523
 DB 1818 ACAGCGCCCCACACAAAGAAATTAATCTGCCCCGAAATGAGTGTGCTGCCAAGCGTAGTA 1877
 OY 3524 ACCTTGTGTTAAAGTAACTGCTGACAGTGTGCTTCCAGAAATTTCCGCTGCTGCTCA 3583
 DB 1878 ACCTTGTGTTAAAGTAACTGCTGACAGTGTGCTTCCAGAAATTTCCGCTGCTGCTCA 1937
 OY 3584 CGTATCATGTTTGAATAATTTTGGCTATTTAAAGATATGATATGATGCTTATTCCTGA 3643
 DB 1938 CGTATCATGTTTGAATAATTTTGGCTATTTAAAGATATGATATGATGCTTATTCCTGA 1997
 OY 3644 TTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3703
 DB 1998 TTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2057
 OY 3704 AGAAGAGGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3763
 DB 2058 AGAAGAGGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2117

OY 3764 TACAAGTATATGAGCTAAATTTTCAGATTTGACAGAGGATTCAGACTTCCTCCATCTCA 3823
 DB 2118 TACAAGTATATGAGCTAAATTTTCAGATTTGACAGAGGATTCAGACTTCCTCCATCTCA 2177
 OY 3824 TGAAGAGAGAGGCTCCCTCATCTGAGCGCTCTGTGAATCTACCTTTGCAAGCTTCAGAC 3883
 DB 2178 TGAAGAGAGAGGCTCCCTCATCTGAGCGCTCTGTGAATCTACCTTTGCAAGCTTCAGAC 2237
 OY 3884 AATACGATGATCTCCCTGAGCAGCAGCGGCTCATCTCTGTGAGGAGGGAAGATTAAGC 3943
 DB 2238 AATACGATGATCTCCCTGAGCAGCAGCGGCTCATCTCTGTGAGGAGGGAAGATTAAGC 2297
 OY 3944 AAAGAGTTAATTTTCAATTCCTCAATCACTAGCTGTAGACTGATCTGTTGTAGCAGTTG 4003
 DB 2298 AAAGAGTTAATTTTCAATTCCTCAATCACTAGCTGTAGACTGATCTGTTGTAGCAGTTG 2357
 OY 4004 TTGTGCTCAATTTTGTGCTGTCGATTTTGTGACATTTTGTGAGANATTTCTATTTGGT 4063
 DB 2358 TTGTGCTCAATTTTGTGCTGTCGATTTTGTGACATTTTGTGAGANATTTCTATTTGGT 2417
 OY 4064 GCTCTACGATTTTGTCTTTTAAATATCTACCTGATATCTGCTTTAAATTTCTTCA 4123
 DB 2418 GCTCTACGATTTTGTCTTTTAAATATCTACCTGATATCTGCTTTAAATTTCTTCA 2477
 OY 4124 CATATGTTTGGCTGATACACTGATTTTATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 4183
 DB 2478 CATATGTTTGGCTGATACACTGATTTTATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 2536
 OY 4184 AAATCTAGTAAGTGCATTTATTTCTTATTAATGACAGCCGTTGCTACTCTGACAGCCT 4243
 DB 2537 AAATCTAGTAAGTGCATTTATTTCTTATTAATGACAGCCGTTGCTACTCTGACAGCCT 2592
 OY 4244 CTCTCTCAATTTTCTTCCGTGACAGTGTGATGCTGATTAACCTATTTTCAATTTGCTTT 4303
 DB 2593 CTCTCTCAATTTTCTTCCGTGACAGTGTGATGCTGATTAACCTATTTTCAATTTGCTTT 2647
 OY 4304 TATTTCTAATATGAGAACATGAGAGTGAACCTTAATTAAGTGTGATTAATAAAT 4363
 DB 2648 TATTTCTAATATGAGAACATGAGAGTGAACCTTAATTAAGTGTGATTAATAAATCT 2707

RESULT 4
 ABV27799
 ID ABV27799 standard; cDNA; 2905 BP.
 XX
 AC ABV27799;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 27790.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200160860-A2.
 PD 23-AUG-2001.
 PF 20-FEB-2001; 2001WO-US05171.
 PR 17-FEB-2000; 2000US-18319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 BA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA Schlegel R, Endege WO, Monahan JE;
 PI

DR WPI: 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1: Page 5714-5715; 11750bp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 2905 BP; 921 A; 509 C; 554 G; 919 T; 2 other;
Query Match 56.1%; Score 2540.4; DB 23; Length 2905;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2627; Conservative 1; Mismatches 62; Indels 10; Gaps 4;
QY 1664 CTGTGAGATTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAATTTCCATCATAAA 1723
DB 18 CCGTGCAGATTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAATTTCCATCATAAA 77
QY 1724 GAGGTACAAAGAGCCAGAACCCAGTATCTTATGCTGGAATGTTAGAACCAAAATCT 1783
DB 78 GAGGTACAAAGAGCCAGAACCCAGTATCTTATGCTGGAATGTTAGAACCCGAATCT 137
QY 1784 CCTAAATTTATTCAGCTCAACAGAGATTTCATCCAGTAATTAATCATCGATCCAGA 1843
DB 138 CCGTAAATTTATTCAGCTCAACAGAGATTTCATCCAGTAATTAATCATCGATCCAGA 197
QY 1844 AGCAGAAAGATTTTAAAGTGGGGAATTAATAAGACCTCATCTTGATTCCTACTGTGC 1903
DB 198 AGCAGAAAGATTTTAAAGTGGGGAATTAATAAGACCTCATCTTGATTCCTACTGTGC 257
QY 1904 AGATTGGGAGCTATTAGTCCCAACATGAAACAGCAAAAGAAATTTTAAAGTGAAGCAG 1963
DB 258 AGATTGGGAGCTATTAGTCCCAACATGAAACAGCAAAAGAAATTTTAAAGTGAAGCAG 317
QY 1964 AAATCTACCTAAAGAGATGTATTCACTGGAATTTATTGGAAGAGATGTTTGTCTACT 2023
DB 318 AAATCTACCTAAAGAGATGTATTCACTGGAATTTATTGGAAGAGATGTTTGTCTACT 377
QY 2024 GTCAACCAAAATATGCTGCAAGTCTTCCAGCCCTGCTGCTTCCAGACACAGAAAGCAA 2083
DB 378 GTCAACCAAAATATGCTGCAAGTCTTCCAGCCCTGCTGCTTCCAGACACAGAAAGCAA 437
QY 2084 AATAGAGAGATCCCATCTACTAGCACATGCAACAGACATAGTTCAAAATTAATACAGA 2143
DB 438 AATAGAGAGATCCCATCTACTAGCACATGCAACAGACATAGTTCAAAATTAATACAGA 497
QY 2144 TGCACCTCTGGAATGTTTCCGGAATACATGTGGAATACTTCCAGTTATTTCACT 2203
DB 498 TGCACCTCTGGAATGTTTCCGGAATACATGTGGAATACTTCCAGTTATTTCACT 557
QY 2204 TCAGAAACCAATTAATGATTTTGTTCAGTGTGACATGTAATCTCAATATAAAAAAGC 2263
DB 558 TCAGAAACCAATTAATGATTTTGTTCAGTGTGACATGTAATCTCAATATAAAAAAGC 617
QY 2264 AATATTGACACTGTGTAAGAGAAATCTTGATTTCACTTACTCATGCTGTTAAATCT 2323
DB 618 AATATTGACACTGTGTAAGAGAAATCTTGATTTCACTTACTCATGCTGTTAAATCT 677

QY 2324 AAGAAATACCTCAGTGGGAGAGAAATCTTCCGGCATATTTTGATCTCTGCCCTCCCT 2383
DB 678 AAGAAATACCTCAGTGGGAGAGAAATCTTCCGGCATATTTTGATCTCTGCCCTCCCT 737
QY 2384 TCCCTCTCTGTTTGTGTAATCTGCATTCAGGTGGCCAGATATTTCCATCTCTTCAGA 2443
DB 738 TCCCTCTCTGTTTGTGTAATCTGCATTCAGGTGGCCAGATATTTCCATCTCTTCAGA 797
QY 2444 CCAGGCTATATTAAGAAACCTTGATTTGGCTCAAGAAATTAAGACAGACTAGA 2503
DB 798 CCAGGCTATATTAAGAAACCTTGATTTGGCTCAAGAAATTAAGACAGACTAGA 857
QY 2504 AAATCATATCACAATTTTACCTGCTCAAGATGGAACCTCTCTTCCAGCTTATGATTT 2563
DB 858 AAATCATATCACAATTTTACCTGCTCAAGATGGAACCTCTCTTCCAGCTTATGATTT 917
QY 2564 TCTAAGTATGATGATGCTCCCAACATCTCAACGCTGACATAGGAAGTCTCCAGTCTAT 2623
DB 918 TCTAAGTATGATGATGCTCCCAACATCTCAACGCTGACATAGGAAGTCTCCAGTCTAT 977
QY 2624 GAAAGAACAGATGTGAGGAGATGATGAAGAACATGAAGATTAATTCGGCAGTCAG 2683
DB 978 GAAAGAACAGATGTGAGGAGATGATGAAGAACATGAAGATTAATTCGGCAGTCAG 1037
QY 2684 AAAAGAACCGATTTGAACCTCTGAGAAATTAAGCATTTGGAATAGAATTTGTTTAAAGA 2743
DB 1038 AAAAGAACCGATTTGAACCTCTGAGAAATTAAGCATTTGGAATAGAATTTGTTTAAAGA 1097
QY 2744 AGCAGAAATATCATTTTATGACGTGATTAAGAGTTTATGATGCTCAAAAGTGAATTTTA 2803
DB 1098 AGCAGAAATATCATTTTATGACGTGATTAAGAGTTTATGATGCTCAAAAGTGAATTTTA 1157
QY 2804 TAGGGCTGTGGTTTCCAAATTTTGGCATGATGATGATTAATTTTCTTAAAGAA 2863
DB 1158 TAGGGCTGTGGTTTCCAAATTTTGGCATGATGATGATTAATTTTCTTAAAGAA 1217
QY 2864 TAATATTAATCATTTTCAAGTTTGCAGATGAGTCCATCCAAATGAATTAATATTAAGT 2923
DB 1218 TAATATTAATCATTTTCAAGTTTGCAGATGAGTCCATCCAAATGAATTAATATTAAGT 1277
QY 2924 CACATATTTTATTTAAATTTTCTAGTACTACTTTAAACAAAGTAAAGTGAAGCAGGCG 2983
DB 1278 CACATATTTTATTTAAATTTTCTAGTACTACTTTAAACAAAGTAAAGTGAAGCAGGCG 1337
QY 2984 AAAATATTTTGAATATGATTTTCAACCCAGTATGATCCCAAAATAGCAGAAATTAAGAA 3043
DB 1338 AAAATATTTTGAATATGATTTTCAACCCAGTATGATCCCAAAATAGCAGAAATTAAGAA 1397
QY 3044 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGAGATACATATTT 3103
DB 1398 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGAGATACATATTT 1457
QY 3104 TTATTAATGAGATATTTTCTTACTTCCGAGTATGACATATTTTCAAGTCTTCTTGCACATG 3163
DB 1458 TTATTAATGAGATATTTTCTTACTTCCGAGTATGACATATTTTCAAGTCTTCTTGCACATG 1517
QY 3164 TGGCCTGTGACTACTGATTTGAGACAGTTCAGTACTAGACAAAATCTGCAATTAATTAATT 3223
DB 1518 TGGCCTGTGACTACTGATTTGAGACAGTTCAGTACTAGACAAAATCTGCAATTAATTAATT 1577
QY 3224 AGTTTACGCCATGATTTCTATTTGGATTTAAATTAATCTAATACAGAGTTAACTGCACA 3283
DB 1578 AGTTTACGCCATGATTTCTATTTGGATTTAAATTAATCTAATACAGAGTTAACTGCACA 1637
QY 3284 GTGCATTCATGACGCTGACAGTATATTTGTTTATTTGAGAGTCAATGATTAATTAATCAGC 3343
DB 1638 GTGCATTCATGACGCTGACAGTATATTTGTTTATTTGAGAGTCAATGATTAATTAATCAGC 1697
QY 3344 GTTGTCAACCTCAGGGGATATTTAGCAATTTGCGGAGACATTTTGTGATGTCAGTACTA 3403
DB 1698 GTTGTCAACCTCAGGGGATATTTAGCAATTTGCGGAGACATTTTGTGATGTCAGTACTA 1757


```

QY 3404 GGGCAGTTATGACATTAGTAGAGAGCCATGAGCTCGCTAAATACCTGCATTCG 3463
    |||
DB 1758 GGGCAGTTATGACATTAGTAGAGAGCCATGAGCTCGCTAAATACCTGCATTCG 1817
QY 3464 ACAGGCCCCCACAACAAGATTATCCCGCAATGCTAGTCCGCAAGGCTGAGTA 3523
    |||
DB 1818 ACAGGCCCCCACAACAAGATTATCCCGCAATGCTAGTCCGCAAGGCTGAGTA 1877
QY 3524 ACCCTGTGTTAAAGTAACCTGAGAGTACCTGCTTCCAGAACTTCCTGCTGCTCA 3583
    |||
DB 1878 ACCCTGTGTTAAAGTAACCTGAGAGTACCTGCTTCCAGAACTTCCTGCTGCTCA 1937
QY 3584 CCGATACCTGTTGAAAAATTTGGCTATTTAAAGATATGATGCTGCTATCTCA 3643
    |||
DB 1938 CCGATACCTGTTGAAAAATTTGGCTATTTAAAGATATGATGCTGCTATCTCA 1997
QY 3644 TTATTACCTGATACACCTGATCTTTCTTAATATTTTCAAGAAAGTGGATTAACCT 3703
    |||
DB 1998 TTATTACCTGATACACCTGATCTTTCTTAATATTTTCAAGAAAGTGGATTAACCT 2057
QY 3704 AGAAGAGCATCAGATGATATTTATTTTAAAGTACCTTAAACCTCTTATTTTC 3763
    |||
DB 2058 AGAAGAGCATCAGATGATATTTATTTTAAAGTACCTTAAACCTCTTATTTTC 2117
QY 3764 TACAAGTTATATGCTAAATTTTCAATGACAGGATTCAGCATTCGCATCTCTCA 3823
    |||
DB 2118 TACAAGTTATATGGTAAATTTTCAATGACAGGATTCAGCATTCGCATCTCTCA 2177
QY 3824 TGGAAAGAGAGGCTCCCTATCTGAAGGCTCTCTGAATCTACCTTCGACAGTTAGAC 3883
    |||
DB 2178 TGGAAAGAGAGGCTCCCTATCTGAAGGCTCTCTGAATCTACCTTCGACAGTTAGAC 2237
QY 3884 AATACGTTGATCTCCCTGAGCAGACAGGCTCATCTGAGGAGGAGGAAAGATTAGCC 3943
    |||
DB 2238 AATACGTTGATCTCCCTGAGCAGACAGGCTCATCTGAGGAGGAGGAAAGATTAGCC 2297
QY 3944 AAAGAGTTAATTTTCAATCAATCACTTACCTGTTAGACTGATCTGTTTGAAGCAGTTG 4003
    |||
DB 2298 AAAGAGTTAATTTTCAATCAATCACTTACCTGTTAGACTGATCTGTTTGAAGCAGTTG 2257
QY 4004 TTGCTCTCATTTTGTCTGCTGATTTTGTGAGACATTTGTTGAGAAATATCTATTTGGT 4063
    |||
DB 2358 TTGCTCTCATTTTGTCTGCTGATTTTGTGAGACATTTGTTGAGAAATATCTATTTGGT 2417
QY 4064 GCTTACTATTTTCTTTTAAATCTACTTATGATATCTGTTTAAATTTCTTCA 4123
    |||
DB 2418 GCTTACTATTTTCTTTTAAATCTACTTATGATATCTGTTTAAATTTCTTCA 2477
QY 4124 CATATGTTTGGCTGATACAGTATTTTATTAAGTAAATTTAAGAAATCTTAACAGCTA 4183
    |||
DB 2478 CATATGTTTGGCTGATACAGTATTTTATTAAGTAAATTTAAGAAATCTTAACAGCTA 2536
QY 4184 AAACCTCAGTAAGTGCATATTTCTTTAAACATAGACCCGTTGCTACCTCAGCCCT 4243
    |||
DB 2537 AAACCTCAGTAAGTGCATATTTCTTTAAATAGAACCGTTGCTACCTCAGCCCT 2592
QY 4244 CTCCTCAATTTTCTTTTCTGATGATGATGCTGATTAACCTCTTCAATTTGCTT 4303
    |||
DB 2593 CTCCTCAATTTTCTTTTCTGATGATGATGCTGATTAACCTCTTCAATTTGCTT 2647
QY 4304 TATTTCTAATATGAGAACATGAGTGAAGTAAATGAGTTGATTAATTAACAT 4363
    |||
DB 2648 TATTTCTAATATGAGAACATGAGTGAAGTAAATGAGTTGATTAATTAACAT 2707

```

```

RESULT 5
AAH05323
ID AAH05323 standard; cDNA; 784 BP.
XX
AC AAH05323;
XX
DT 26-JUN-2001 (first entry)
XX

```

```

DE Human cDNA clone (5'-primer) SEQ ID NO:2158.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 2158; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 784 BP; 262 A; 149 C; 148 G; 222 T; 3 other;
XX
Query Match 17.1%; Score 776.2; DB 22; Length 784;
Best Local Similarity 99.2%; Pred. No. 2.9e-170;
Matches 778; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1686 TGACCTAACCAAAATGTTCTGTAATTTCCATCATTAAGTACCAAGAGCGGAGAC 1745
    |||
DB 1 TGTACTAACCAAAATGTTCTGTAATTTCCATCATTAAGTACCAAGAGCGGAGAC 60
QY 1746 CCAGTATCTTATGCTGGAATGTTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 1805
    |||
DB 61 CCAGTATCTTATGCTGGAATGTTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 120
QY 1806 AGGATTTCAATCCAGTGAATATACATGATCCAAAGAACAGAGATATTTAAGTGGG 1865
    |||
DB 121 AGGATTTCAATCCAGTGAATATACATGATCCAAAGAACAGAGATATTTAAGTGGG 180

```

```
QY 1866 GAATTATATAAGACCTCATCTTGTATCTAGTGTCTGACATTTGGACTATTTAGTCCA 1925
DB 181 GAAATTATATAAGACCTCATCTTGTATCTAGTGTCTGACATTTGGACTATTTAGTCCA 240
QY 1926 ACCATGAAACAGCAAGAAAGATTTAGTGAAGCAGAAACCTCAAAAGGATATGTT 1985
DB 241 ACCATGAAACAGCAAGAAAGATTTAGTGAAGCAGAAACCTCAAAAGGATATGTT 300
QY 1986 ATCACTGGAATTTATCTGAAAGATGTTTTCCTACTGTCACCAATATGTGCAAGT 2045
DB 301 ATCACTGGAATTTATCTGAAAGATGTTTTCCTACTGTCACCAATATGTGCAAGT 360
QY 2046 CTTCACACCTCTGCTGCTGCCAGACACACAGAAAGCAAAATAGAGACATCCACTAGCT 2105
DB 361 CTTCACACCTCTGCTGCTGCCAGACACACAGAAAGCAAAATAGAGACATCCACTAGCT 420
QY 2106 AGCACACATGCACAGACATAGTTCAATATATACAGATGCATCTAGTGAATGTTCCG 2165
DB 421 AGCACACATGCACAGACATAGTTCAATATATACAGATGCATCTAGTGAATGTTCCG 480
QY 2166 GAAATCTGTGGAATAATCTTCCAGTATTTTCAGACTTCAGAAACCATTTATGATTTG 2225
DB 481 GAAATCTGTGGAATAATCTTCCAGTATTTTCAGACTTCAGAAACCATTTATGATTTG 540
QY 2226 TTCACTGATGGCACTGTAATCTCTCAATATATATAAAACCAATATTGACACTGTGAACAG 2285
DB 541 TTCACTGATGGCACTGTAATCTCTCAATATATAAAACCAATATTGACACTGTGAACAG 600
QY 2286 AAATCTGTGATCATTTACTCCATGCTGTTAAATCTAATAAGAAATCTCCAGTGGGAGA 2345
DB 601 AAATCTGTGATCATTTACTCCATGCTGTTAAATCTAATAAGAAATCTCCAGTGGGAGA 660
QY 2346 GGAATCTGTGGGCATATTTGATCTCTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 2405
DB 661 GGAATCTGTGGGCATATTTGATCTCTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
QY 2406 CTGCACTTCAGTGGCCAGATTTGCTCTTCCTTCAGACACAGGCTATATTTGAAGAAAC 2465
DB 721 CTGCACTTCAGTGGCCAGATTTGCTCTTCCTTCAGACACAGGCTATATTTGAAGAAAC 780
QY 2466 CTTG 2469
DB 781 CTTG 784

RESULT 6
ABV12212
ID ABV12212 standard; cDNA; 360 BP.
XX
AC ABV12212;
XX
XX 13-SEP-2002 (first entry)
DT
XX
DE Human prostate expression marker cDNA 12203.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
XX
PN WO200106860-A2.
PD
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
```

```
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
PS Claim 1; Page 2006; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 360 BP; 124 A; 68 C; 73 G; 94 T; 1 other;
XX
Query Match 7.7%; Score 349.8; DB 23; Length 360;
Best Local Similarity 99.4%; Pred. No. 2.5e-71;
Matches 351; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1725 ATGTACAGAAAGCGGAGAACCCAGTATCTTATGCGAATGTTAGAACCAAGATCTC 1784
DB 7 AGGTACAGAAAGCGGAGAACCCAGTATCTTATGCGAATGTTAGAACCAAGATCTC 66
QY 1785 CTAAAAATTATCCAGCTCAACAGAGATTTCATATCAGTGAATATTAACATGATCCAGAA 1844
DB 67 CTAAAAATTATCCAGCTCAACAGAGATTTCATATCAGTGAATATTAACATGATCCAGAA 126
QY 1845 GCAGAAAGATTTAAGTGGGGAATATATTAAGACCTCATCTTGTATCTATGTGTC 1904
DB 127 GCAGAAAGATTTAAGTGGGGAATATATTAAGACCTCATCTTGTATCTATGTGTC 186
QY 1905 GATTTGGACTATTTAGTCCCAACCATGAAACAGCAAAAGATTTAGTGAAGCAGA 1964
DB 187 GATTTGGACTATTTAGTCCCAACCATGAAACAGCAAAAGATTTAGTGAAGCAGA 246
QY 1965 AACTACCTAAAGGATATGTTATCTACTGCAATTTATCTGAAGAGATGTTTGTACTG 2024
DB 247 AACTACCTAAAGGATATGTTATCTACTGCAATTTATCTGAAGAGATGTTTGTACTG 306
QY 2025 TCACCAAAATATGCTCAAGTCTTCAGCCCTCTGCTCCGACACACAGA 2077
DB 307 TCACCAAAATATGCTCAAGTCTTCAGCCCTCTGCTCCGACACACAGA 359

RESULT 7
ABV03043
ID ABV03043 standard; cDNA; 374 BP.
XX
AC ABV03043;
XX
XX 13-SEP-2002 (first entry)
DT
XX
DE Human prostate expression marker cDNA 3074.
XX
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
```



```

XX  WO200160860-A2.
PN
XX  23-AUG-2001.
PD
XX  20-FEB-2001; 2001WO-US05171.
PF
XX  17-FEB-2000; 2000US-183319P.
PR  16-MAR-2000; 2000US-189862P.
PR  25-MAY-2000; 2000US-207454P.
PR  09-JUN-2000; 2000US-211314P.
PR  18-JUL-2000; 2000US-219007P.
PR  13-DEC-2000; 2000US-255281P.
XX
XX  (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX  Schlegel R, Endege WO, Monahan JE;
XX
XX  WPI; 2001-662795/76.
XX
XX  Novel isolated nucleic acid molecule associated with cancerous state of
XX  prostate cells and correlating with presence of prostate cancer, useful
XX  for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX  Claim 1; Page 561; 11750pp; English.
XX
XX  The invention relates to an isolated nucleic acid molecule (I) comprising
XX  a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX  specification or its complement. (I) is useful for:
XX  (a) assessing whether a patient is afflicted with prostate cancer;
XX  (b) monitoring the progression of prostate cancer in a patient;
XX  (c) assessing the efficacy of a test compound to inhibit prostate
XX  cancer in a patient;
XX  (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX  in a patient;
XX  (e) selecting a composition for inhibiting prostate cancer in a patient;
XX  (f) assessing the prostate cell carcinogenic potential of a compound;
XX  (g) determining whether prostate cancer has metastasized in a patient;
XX  (h) assessing the aggressiveness or indolence of prostate cancer in a
XX  patient;
XX  (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX  Sequence 374 BP; 123 A; 75 C; 81 G; 95 T; 0 other;
SQ
XX
XX  Query Match      7.7%; Score 348.8; DB 23; Length 374;
XX  Best Local Similarity 99.4%; Pred. No. 4.3e-71;
XX  Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
OY  1725 ATGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 1784
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  23 AGGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 82
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  1785 CTAATAATTATTCACGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 1844
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  83 CTAATAATTATTCACGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 142
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  1845 GCAGAGAATATTTAAGTGGGGAATTTATTAAGAACCTCATCTTGATCTGTGTGTCA 1904
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  143 GCAGAGAATATTTAAGTGGGGAATTTATTAAGAACCTCATCTTGATCTGTGTGTCA 202
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  1905 GATTGGAGCTATTAGTCCACATGCAAAACAGCAAAAGAGATTTTGTAGACAGCA 1964
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  203 GATTGGAGCTATTAGTCCACATGCAAAACAGCAAAAGAGATTTTGTAGACAGCA 262
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  1965 AACTACCTAAAGAGATATGTTATTCACCTGGAATTTATTTCTGAAGAGATGTTTGTACTG 2024
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  263 AACTACCTAAAGAGATATGTTATTCACCTGGAATTTATTTCTGAAGAGATGTTTGTACTG 322
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  2025 TCAACCAATATGTCGCAAGTCTTCCAGCCCTGCTGCTTGCAGACACAG 2076
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  323 TCAACCAATATGTCGCAAGTCTTCCAGCCCTGCTGCTTGCAGACACAG 374
XX

```

```

RESULT 8
ABV33357
ID  ABV33357 standard; cDNA; 397 BP.
XX
XX  ABV33357;
XX
XX  16-SEP-2002 (first entry)
XX
XX  Human prostate expression marker cDNA 33348.
XX
XX  Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX  pharmacogenomic marker; gene; ss.
XX
XX  Homo sapiens.
XX
XX  WO200160860-A2.
XX
XX  23-AUG-2001.
XX
XX  20-FEB-2001; 2001WO-US05171.
XX
XX  17-FEB-2000; 2000US-183319P.
XX  16-MAR-2000; 2000US-189862P.
XX  25-MAY-2000; 2000US-207454P.
XX  09-JUN-2000; 2000US-211314P.
XX  18-JUL-2000; 2000US-219007P.
XX  13-DEC-2000; 2000US-255281P.
XX
XX  (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX  Schlegel R, Endege WO, Monahan JE;
XX
XX  WPI; 2001-662795/76.
XX
XX  Novel isolated nucleic acid molecule associated with cancerous state of
XX  prostate cells and correlating with presence of prostate cancer, useful
XX  for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX  Claim 1; Page 7066; 11750pp; English.
XX
XX  The invention relates to an isolated nucleic acid molecule (I) comprising
XX  a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX  specification or its complement. (I) is useful for:
XX  (a) assessing whether a patient is afflicted with prostate cancer;
XX  (b) monitoring the progression of prostate cancer in a patient;
XX  (c) assessing the efficacy of a test compound to inhibit prostate
XX  cancer in a patient;
XX  (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX  in a patient;
XX  (e) selecting a composition for inhibiting prostate cancer in a patient;
XX  (f) assessing the prostate cell carcinogenic potential of a compound;
XX  (g) determining whether prostate cancer has metastasized in a patient;
XX  (h) assessing the aggressiveness or indolence of prostate cancer in a
XX  patient;
XX  (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX  Sequence 397 BP; 129 A; 80 C; 88 G; 100 T; 0 other;
SQ
XX
XX  Query Match      7.7%; Score 348.8; DB 23; Length 397;
XX  Best Local Similarity 99.4%; Pred. No. 4.4e-71;
XX  Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
OY  1725 ATGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 1784
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  46 AGGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 105
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  1785 CTAATAATTATTCACGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 1844
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  106 CTAATAATTATTCACGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 165
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  1845 GCAGAGAATATTTAAGTGGGGAATTTATTAAGAACCTCATCTTGATCTGTGTGTCA 1904
XX  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX  166 GCAGAGAATATTTAAGTGGGGAATTTATTAAGAACCTCATCTTGATCTGTGTGTCA 225
XX

```

QY 1905 GTATTGGACTATTAGTCCACCATGAAAAAGAGATTTTGTGAGCAGCA 1964
|||||
DB 226 GTATTGGACTATTAGTCCACCATGAAAAAGAGATTTTGTGAGCAGCA 285
|||||
QY 1965 AACTACCTAAAGATATGTTATCTGGAATTTATCTGAGAAGATGTTTGTCTACTG 2024
|||||
DB 286 AACTACCTAAAGATATGTTATCTGGAATTTATCTGAGAAGATGTTTGTCTACTG 345
|||||
QY 2025 TCACCAAAATATGCTGCAAGCTTCCAGCCCTGCTGCTGCAGACACACAG 2076
|||||
DB 346 TCACCAAAATATGCTGCAAGCTTCCAGCCCTGCTGCTGCAGACACACAG 397
|||||
RESULT 9
ABV42280
ID ABV42280 standard; cDNA; 397 BP.
XX
AC ABV42280:
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 42271.
XX
KM Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-ANG-2001.
XX
PF 20-FEB-2001; 2001MO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX
PS Claim 1; Page 8470-8471; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
SQ Sequence 397 BP; 129 A; 80 C; 88 G; 100 T; 0 other;

Query Match

7.7%; Score 348.8; DB 23; Length 397;

Best Local Similarity 99.4%; Pred. No. 4,4e-71;
Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1725 ATGTACAGAGAAAGCCGAGAACCCAGTATCTTATGCTGGAATGTTAGGAACCAAGATCTC 1784
|||||
DB 46 AGGTACAGAGAAAGCCGAGAACCCAGTATCTTATGCTGGAATGTTAGGAACCCAGATCTC 105
|||||
QY 1785 CTAAATTTATCCAGCTCAACAGAGATTTCAATTCAGTGAATTAACATCATCAAGAA 1844
|||||
DB 106 CTAAATTTATCCAGCTCAACAGAGATTTCAATTCAGTGAATTAACATCATCAAGAA 165
|||||
QY 1845 GCAGAGAGATTTTAAAGTGGGGAATTTATATAAGACCTCATCTTGTATTTAGTGTCTCA 1904
|||||
DB 166 GCAGAGAGATTTTAAAGTGGGGAATTTATATAAGACCTCATCTTGTATTTAGTGTCTCA 225
|||||
QY 1905 GTATTGGACTATTAGTCCACCATGAAAAAGAGATTTTGTGAGCAGCA 1964
|||||
DB 226 GTATTGGACTATTAGTCCACCATGAAAAAGAGATTTTGTGAGCAGCA 285
|||||
QY 1965 AACTACCTAAAGATATGTTATCTGGAATTTATCTGAGAAGATGTTTGTCTACTG 2024
|||||
DB 286 AACTACCTAAAGATATGTTATCTGGAATTTATCTGAGAAGATGTTTGTCTACTG 345
|||||
QY 2025 TCACCAAAATATGCTGCAAGCTTCCAGCCCTGCTGCTGCAGACACACAG 2076
|||||
DB 346 TCACCAAAATATGCTGCAAGCTTCCAGCCCTGCTGCTGCAGACACACAG 397
|||||
RESULT 10
AAC28836
ID AAC28836 standard; cDNA; 355 BP.
XX
AC AAC28836:
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 32911.
XX
KM Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GENE) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX
PS Claim 1; SEQ ID 32911; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

50 Sequence 355 BP; 133 A; 50 C; 72 G; 94 T; 6 other;

Query Match	7.5%	Score 341	DB 21	Length 355
Best Local Similarity	98.3%	Pred. No. 2.7e-69		
Matches 349	Conservative 5	Mismatches 0	Indels 1	Gaps 1

QY	2513	CACAAATTTTACGCGTCACAAAGTGGAAACCTCCTCTCCAGCTTTATGATTTTCTTAAGTAT	2572
Db	1	CACAAATTTTACCTGSSKCAAAAGATGGAAACCTCCTCTCCAGCTTTATGATTTTCTTAAGTAT	60
QY	2573	GATAGATGCGCGCAACATCTCAACGTGGCACATAGGAAAGTTTCCCAAGTATGTAAGAAAGAAC	2632
Db	61	GATAGATGCGCGCAACATCTCAACGTGGCACATAGGAAAGTTTCCCAAGTATGTAAGAAAGAAC	120
QY	2633	AGATGTGCAGAGAAATGATTAAGGAAACACATGCAAGTAAATCGGCAGTCAGAAAAAGAAC	2692
Db	121	AGATGTGCAGAGAAATGATTAAGGAAACACATGCAAGTAAATCGGCAGTCAGAAAAAGAAC	180
QY	2693	GATTGAACAATCTGAGAAATAAAGCATTTGGAATAGAAATAATTGTTTAAAGAACGAGAAAA	2752
Db	181	GATTGAACAATCTGAGAAATAAAGCATTTGGAATAGAAATAATTGTTTAAAGAACGAGAAAA	240
QY	2753	ATCATTTTAGACGTGATTAAGAGATTAGCATGCTCAAAAGTGAACATAATTTTATAGGCGTGT	2812
Db	241	ATCATTTTAGACGTGATTAAGAGATTAGCATGCTCAAAAGTGAACATAATTTTATAGGCGTGT	300
QY	2813	GGTTTCCAAATTTTTTTGGCAGTA-TAGACTAATTTTATTTCCCTTAAGAAATA	2866
Db	301	GGTTTCCAAATTTTTTTGGCAGTAAATTAATTTTCCCTTAAGAAATA	355

RESULT 11
ABV38274
ID ABV38274 standard; cDNA; 416 BP.

AC	ABV38274;	
XX		
DT	16-SEP-2002	(first entry)
XX		
DE	Human prostate expression marker cDNA	38265

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker
 KW pharmacogenomic marker; gene; ss.

Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.

PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX	Claim 1; Page 7806-7807; 11750pp; English.
PS	
...	

CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

SQ Sequence 416 BP; 111 A; 78 C; 66 G; 161 T; 0 other;

Query Match:	6.8%:	Score 307.6:	DB 23:	Length 416:
Best Local Similarity:	98.1%:	Pred. No. 1,7e-61:		
Matches 310:	Conservative 1:	Mismatches 5:	Indels 0:	Gaps 0:

OY	4090	TCACCTGATATCTGTTCTTTAAATTTTCTACATATGGTTGGCATCAACATAT	4149
Db	21	TCACCTGATATCTGTTCTTTAAATTTTTCACATATGGTTGGCATCAACATAT	80
OY	4150	TTTTATACGTGAATTTTAAAGAAATCTACACGCTAAACCTCAGTAAGTCATWTATTCCT	4209
Db	81	TTTTATACGTGAATTTTAAAGAAATCTACACGCTAAACCTCAGTAAGTCATWTATTCCT	140
OY	4210	TATACATTAACCCGTTGGTACCTCTAGCACCCTCCCTCATATTTTTCCTGTAGCAT	4269
Db	141	TATACATTAACCCGTTGGTACCTCTAGCACCCTCCCTCATATTTTTCCTGTAGCAT	200
OY	4270	GTCATGCGCTGATTAACATCTTTTCATTTTGCTTTTATTTCTAATATGGAACATGAGAG	4329
Db	201	GTCATGCGCTGATTAACATCTTTTCATTTTGCTTTTATTTCTAATATGGAACAAAGAGAG	260
OY	4330	TGAACCTAAATATAGGTTTGTAGTATTAACATCATCTATGAGCTAATTTATGAAATAGCT	4389
Db	261	TGAACCTAAATATAGGTTTGTAGTATTAACATCATCTATGAGCTAATTTATGAAATAGCT	320
OY	4390	AATTAAGTACAGCAGC 4405	
Db	321	AATTAAGTACCTGCC 336	

RESULT 12
ABV25366
ID ABV25366 standard; cDNA; 416 BP

AC ABV25366;

DT 16-SEP-2002 (first entry)

DE Human prostate expression marker CDNA 25357.

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker
KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.
vxy

PR 17-FEB-2000; 2000US-183319P,
PR 16-MAR-2000; 2000US-189862P,

PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX
XX Claim 1; Page 5006; 11750pp; English.
XX
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX
XX Sequence 416 BP; 109 A; 67 C; 57 G; 159 T; 24 other;
SQ
Query Match 6.7%; Score 304.4; DB 23; Length 416;
Best Local Similarity 87.1%; Pred. No. 9, 4e-61;
Matches 317; Conservative 1; Mismatches 46; Indels 0; Gaps 0;
QY 4090 TCTACTGATATCTGCTCTTAAATTTCTTCACATATGTTGGCTGATACACTGAT 4149
DB 21 TCTACTGATATCTGCTCTTAAATTTCTTCACATATGTTGGCTGATACACTGAT 80
QY 4150 TTTTAACTGAATTAAGATCTAAGATCTAAGATCTAAGATCTAAGATCTAAGATCT 4209
DB 81 TTTTAACTGAATTAAGATCTAAGATCTAAGATCTAAGATCTAAGATCTAAGATCT 140
QY 4210 TTTAATCATAGACCGCTTGTCTACCTCAGACCGCTTGTCTACCTCAGACCGCTTGT 4269
DB 141 TTTAATCATAGACCGCTTGTCTACCTCAGACCGCTTGTCTACCTCAGACCGCTTGT 200
QY 4270 GTGATCGCTGATTAACATCTATTTGCTTTTCTTAAATTTCTTAAATTTCTTAAATTT 4329
DB 201 GTGATCGCTGATTAACATCTATTTGCTTTTCTTAAATTTCTTAAATTTCTTAAATTT 260
QY 4330 TGAATCTTAATATAGTGTGTAGTATTAATAACATCTATTTGCTTTTCTTAAATTT 4389
DB 261 TGAATCTTAATATAGTGTGTAGTATTAATAACATCTATTTGCTTTTCTTAAATTT 320
QY 4390 AATTAGTACGACGACATGAATGAATGCTTACGATGTACCTTTGTCACACAT 4449
DB 321 AATTAGTACGACGACATGAATGAATGCTTACGATGTACCTTTGTCACACAT 380
QY 4450 TTTG 4453
DB 381 TATG 384
RESULT 13
ID AAF93519 standard; cDNA; 547 BP.
XX AAF93519;
AC

XX
XX 21-MAY-2001 (first entry)
XX
XX cDNA encoding SRT protein isolated from human retina tissue SEQ ID 340.
DE
XX Human; SRT; gene therapy; gene mapping; tissue typing; ss.
XX
XX Homo sapiens.
XX
XX WO200107611-A2.
XX
XX
XX 01-FEB-2001.
XX
XX 21-JUL-2000; 2000WO-US20006.
XX
XX 26-JUL-1999; 99US-0145701.
XX
XX
XX (GEN) GENENTECH INC.
XX
XX Baker KP, Goddard A, Wood WT;
XX WPI; 2001-112729/12.
XX
XX
XX New isolated nucleic acid molecule encoding a SRT polypeptide is useful
PT for production of recombinant SRT polypeptides, gene mapping,
PT diagnosing genetic disorders and for gene therapy -
XX
XX
XX Claim 2; Fig 340; 663pp; English.
XX
XX
XX Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
CC human SRT proteins. The cDNA sequences are isolated from various
CC different human tissue cDNA libraries. The invention relates to a method
CC for detecting cDNA encoding an SRT protein, a vector containing cDNA
CC encoding SRT, a host cell transformed with the vector, an isolated SRT
CC polypeptide, and an antibody which binds to SRT. The polynucleotide
CC sequence can be used in gene therapy and is useful in the recombinant
CC production of SRT polypeptides, as a hybridisation probe to screen
CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
CC map the gene encoding the SRT polypeptides and analysing genetic
CC disorders, tissue typing and disease tissue detection. The SRT
CC polynucleotide sequences can be used in polymerase chain reaction,
CC screening for new therapeutic molecules and generation of antisense RNA
CC and DNA.
XX
XX
XX Sequence 547 BP; 155 A; 117 C; 131 G; 139 T; 5 other;
SQ
Query Match 6.4%; Score 290; DB 22; Length 547;
Best Local Similarity 88.8%; Pred. No. 2, 3e-57;
Matches 348; Conservative 0; Mismatches 37; Indels 7; Gaps 3;
QY 111 CCTCTCCGTCGCGTGGCGGGAACCTGTGACCGCGCGCTCGGGAACGCCAGCT 170
DB 114 CCGCTCCGTCGCGTGGCGGGAACCTGTGACCGCGCGCTCGGGAACGCCAGCT 172
QY 171 CCGCGCGCGTGGCGGGAACCTGTGACCGCGCGCTCGGGAACGCCAGCT 230
DB 173 CCGCGCGCGTGGCGGGAACCTGTGACCGCGCGCTCGGGAACGCCAGCT 228
QY 231 TTCAGAAAATTTGAATAGCAAAATAGAAATTAAGAAATTAAGCAAGATACAGAG 290
DB 229 TTCAGAAAATTTGAATAGCAAAATAGAAATTAAGAAATTAAGCAAGATACAGAG 288
QY 291 ACAGATGGAAGTGTCTTGAAGAAACAGACAGCAGTGAAGAAACAGACAAATCC 350
DB 289 --CAGCATGAAGTGTCTTGAAGAAACAGACAGCAGTGAAGAAACAGACAAATCC 346
QY 351 GCTCAGATCAACAGTGTATGTTTCCGCTTCAATGCTTTAGAGTTGGAGTC 410
DB 347 GCTCAGATCAACAGTGTATGTTTCCGCTTCAATGCTTTAGAGTTGGAGTC 406
QY 411 TCTTTTGCATATGTCATTTTAAATGCAACAGTAACTCTTTACAGACAGT 470
DB 407 TCTTTTGCATATGTCATTTTAAATGCAACAGTAACTCTTTACAGACAGT 466

OY 471 CCTCAGAAATATTTAGTACATTGCACAGG 502
|||||
DB 467 CCTCAGAAATATTTAGTACATTGCACAGG 498

RESULT 14
ABV08367
ID ABV08367 standard; cDNA: 447 BP.
XX
AC ABV08367;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 8358.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI: 2001-662795/76.
XX
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 1328; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX
SQ Sequence 447 BP; 116 A; 86 C; 75 G; 168 T; 2 other;

Query Match 5.2%; Score 234.6; DB 23; Length 447;
Best Local Similarity 92.6%; Pred. No. 1.6e-44;
Matches 301; Conservative 1; Mismatches 15; Indels 8; Gaps 5;

OY 4089 ATCTACTGATATCTTGTCTTAATTTCTTCACATATGGTTGCCGATACAACTGA 4148
|||||
DB 43 ATCTACTGATATCTTGTCTTAAATTTTTCACATATGGTTGCCGATACAACTGA 102
|||||

OY 4149 TTTTATTAAGTAATTTAAGGAATCTAACAGCTAAACCTCAGTAAGTCATMTATTCC 4208
|||||

DB 103 TTTTATTAAGTAATTTAAGGAATCTAACAGCTAAGTCATATATCTCC 162
|||||

OY 4209 TTATAA-CATGAGACCGGTGGTCTACTCTC-AGACACCTCTCCCAATTTTTTCTGTGAG 4266
|||||

DB 163 TTATTAACCATAGACCCGCGGTACTCTCAAGCACCCCTCCCAATTTTTTCCGTGAG 222
|||||

OY 4267 CATGTGATGCCGATTAAACCATTTTCA-TTGTCTTATTTCAATATATGAGG---AAC 4321
|||||

DB 223 CATGTGATGCCGATTAAACCATTTTCA-TTGTCTTATTTCAATATATGAGGGAACA 282
|||||

OY 4322 AATGAGAGTGAACCTCAATATATA-GTTGTAGTAAATAAACATCAATAGCTAATATTA 4380
|||||

DB 283 TTGAGAAAGTGAACCTCAATATATAGGTTGTAGTAATAAACATCAATAGCTAATATTA 342
|||||

OY 4381 GAAATGCTAATTAATGATACACAC 4405
|||||

DB 343 GAAATGCTAATTAATGATACCTGCC 367
|||||

RESULT 15
AAS26179
ID AAS26179 standard; cDNA: 1423 BP.
XX
AC AAS26179;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 358.
XX
KW Human; immunosuppressive; antiarthritic; ss; antihemetic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnery; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.

XX
OS Homo sapiens.
XX
PN WO20015322-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-023142.
 PR 08-SEP-2000; 2000US-0231423.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235634.
 PR 27-SEP-2000; 2000US-0235636.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249399.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI: 2001-488783/53.
 DR P-PSDB: AAU16192.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; SEQ ID NO 358; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g.
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage

CC capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present CC sequence encodes a novel secreted protein of the invention.

Query Match 2.38; Score 106; DB 22; Length 1423;

Best Local Similarity 63.28; Pred. No. 2e-14;

Matches 230; Conservative 0; Mismatches 125; Indels 9; Gaps 4;

```
QY 2844 TAATTTATTTCCCTAAAGCATATATTTAATCATTTCAAGTTGCGAGCTAGTGCATCC 2903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 TAACTTTGGCTACTACTCAATATATACCACTTAATATATATGTTCTAGAGCAGCGCTGCC 710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2904 AATAGAAATTAATAATATAGTCACATATTTTAAATTTCTAGTAACATTAAC 2963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 AATAGAAATTAATACTGAGCCACATGTAATTTT--ATTTCTCTAGCCACATTAATA 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2964 AAAGTAAAGTGACAGCGCAAAA-----TAATTTGATATTACTTTTCCACCAGTAGTA 3018
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AAGTAAAGATACAGACAGTAATTTTAAATGTTTAAATTCAGTATATCCAAATATG 828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3019 TACCCAAATATGCGAAATATATATATATATATGATATTTTACATCCTTTT--TTGTAC 3077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 ATTGAACATCTAATATATATATAAATTTATATGATATTTTACATCTTTTGTATA 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3078 CAAGTCTTCTAATATGACATATTTTATATCTTACTGCAATTTCTTACTCCGAGTAGCC 3137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 CTAGTCTTCAAAATCTGATGATCTTACATGATGATAGCACATCTCACTTTGTACTAGCC 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3138 ATATTTCAGTGTCA--TTGCCACATGCGCTGAGTACTGATTTGGACAGTTCAGTA 3196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 ACATTGCAAGTGTGACAGTACATGCGCTAGTGGCTACTGCACTGACAGCACAGTT 1008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3197 CTAG 3200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1009 CTAG 1012
```

Search completed: March 22, 2003, 08:54:57
Job time : 1003 secs

THIS PAGE BLANK (USPTO)

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Run on: March 22, 2003, 03:03:41 ; Search time 12776 Seconds

(without alignments)
10309.901 Million cell updates/sec

Title: US-09-847-046-1

Sequence: 1 gccgagtcgtggcgctgc.....ccttcaaaaaaaaaaaaaa 4526

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
```

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Maximum Match 100%
Listing first 45 summaries

Database :

1: gb_bmb1:*
2: gb_hlg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_v1:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_v1:*
30: em_hlg_hum:*
31: em_hlg_in:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pin:*
35: em_hlg_rod:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	4433.6	98.0	4552	9	AB037765	AB037765 Homo sapi
2	2811.6	62.1	2820	9	AK022563	AK022563 Homo sapi
3	2410.2	53.3	2747	9	AK097748	AK097748 Homo sapi
4	1995.8	44.1	161100	2	AC025659	AC025659 Homo sapi
5	1985.8	44.1	176368	9	CNS01RGG	AL157971 Human chr
6	426.4	9.4	1496	10	BC027108	BC027108 Mus muscu
7	307.2	6.8	204606	2	AC021131	AC021131 Homo sapi
8	290	5.4	547	6	AX079596	AX079596 Sequence
9	233.6	5.2	204606	2	AC021131	AC021131 Homo sapi
10	201.4	4.4	212103	9	CNS01DVA	AL133453 Human chr
11	129.4	2.9	203610	2	AC102579	AC102579 Mus muscu
12	126.6	2.8	203610	2	AC102579	AC102579 Mus muscu
13	124	2.7	94749	9	HS503811	AL031122 Human DNA
14	115	2.5	14499	9	AC094089	AC094089 Homo sapi
15	111	2.5	89196	9	AC012619	AC012619 Homo sapi
16	111	2.5	100457	2	AC020933	AC020933 Homo sapi
17	111	2.5	123641	2	AC012605	AC012605 Homo sapi
18	111	2.5	154054	2	AC020919	AC020919 Homo sapi
19	111	2.5	155766	2	AC114961	AC114961 Homo sapi
20	111	2.5	170970	2	AC021564	AC021564 Homo sapi
21	111	2.5	264095	9	AC008050	AC008050 Homo sapi
22	110.2	2.4	172388	2	AC027634	AC027634 Homo sapi
23	110.2	2.4	176814	9	AC100849	AC100849 Homo sapi
24	109.2	2.4	174710	2	AC069134	AC069134 Homo sapi
25	109.2	2.4	176807	9	HS198P4	AC037487 Homo sapi
26	108.4	2.4	35714	9	CNS07EFA	AL008708 Human DNA
27	107.2	2.4	158305	9	AC068794	AL151231 Human chr
28	106	2.3	137043	9	AC126470	AC126470 Homo sapi
29	105.6	2.3	196710	2	AC019054	AC019054 Homo sapi
30	105.6	2.3	169845	2	AC091045	AC091045 Homo sapi
31	105.6	2.3	171947	9	AC021810	AC021810 Homo sapi
32	105.6	2.3	192100	9	AC007628	AC007628 Genomic S
33	105.4	2.3	140356	9	AC090208	AC090208 Homo sapi
34	105.4	2.3	200829	2	AC130197	AC130197 Fells cat
35	105.2	2.3	158690	2	AC0944102	AC0944102 Homo sapi
36	105	2.3	118985	9	AC024652	AC024652 Homo sapi
37	105	2.3	176552	2	AC116362	AC116362 Homo sapi
38	105	2.3	178530	9	AC023479	AC023479 Homo sapi
39	104.8	2.3	166258	2	AL136231	AL136231 Human DNA
40	104.8	2.3	160012	2	AP002405	AP002405 Homo sapi
41	104.6	2.3	167886	2	AC097466	AC097466 Homo sapi
42	104.6	2.3	170321	2	AC011777	AC011777 Homo sapi
43	104.4	2.3	184332	9	AC002851	AC002851 Homo sapi
44	104.4	2.3	122674	9	AC007021	AC007021 Homo sapi
45	103.8	2.3	200610	9	AC007021	AC007021 Homo sapi

ALIGNMENTS

RESULT	1
AB037765	
LOCUS	AB037765 4552 bp mRNA linear PRI 10-MAY-2002
DEFINITION	Homo_sapiens mRNA for KIAA1344 protein, partial cds.
ACCESSION	AB037765
VERSION	AB037765.2 GI:20521893
KEYWORDS	.
SOURCE	Homo sapiens brain cDNA to mRNA, clone_lib:pblunescriptII SK plus clone:fj00476sl. Homo sapiens
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 Nagase,T., kikuno,R., Ishikawa,K.I., Hirosewa,M. and Ohara,O.
TITLE	Prediction of the coding sequences of unidentified human genes.

|||||
Db 1321 TGGGGCTTCAAAAGAGCAGAGAGGAGCTCCAGTGAATTTTGTATTCATGATGT 1380
Oy 1340 TGAATTAATATATCATGATGGAATATATATGACATTTGGGAATACAGAAATGA 1399
Db 1381 TGAATTAATTAATATATCATGATGGAATATATATGACATTTGGGAATACAGAAATGA 1440
Oy 1400 AGACATATGATGAGAGGTCAGATATAGATGTTCCAGATGATGAGATGAGAACTGT 1459
Db 1441 AGACATATGATGAGAGGTCAGATATAGATGTTCCAGATGATGAGATGAGAACTGT 1500
Oy 1460 TTTCCAGAGATGAGAGAGAAATTAACCTTTGGAATTAAGTGAATTAACAGAGAAAC 1519
Db 1501 TTTCCAGAGATGAGAGAGAAATTAACCTTTGGAATTAAGTGAATTAACAGAGAAAC 1560
Oy 1520 ATTTAATGACAGATGAGAGGCTTCAGAGCATGATGATGATGATGATGATGATGATG 1579
Db 1561 ATTTAATGACAGATGAGAGGCTTCAGAGCATGATGATGATGATGATGATGATGATG 1620
Oy 1580 AGTATCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1639
Db 1621 AGTATCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
Oy 1640 TACTATGCTTCTTACTAGATTAACCTGTCAGATGATGATGATGATGATGATGATGATG 1699
Db 1681 TACTATGCTTCTTACTAGATTAACCTGTCAGATGATGATGATGATGATGATGATGATG 1740
Oy 1700 TGTACTGATATTTCTATCATTAAGATGATGATGATGATGATGATGATGATGATGATG 1759
Db 1741 TGTACTGATATTTCTATCATTAAGATGATGATGATGATGATGATGATGATGATGATG 1800
Oy 1760 TGAATGTTAGAGAACCAAGATCTCTTAATTTATCCAGTCAACAGATTTGATATCC 1819
Db 1801 TGAATGTTAGAGAACCAAGATCTCTTAATTTATCCAGTCAACAGATTTGATATCC 1860
Oy 1820 AGTGAATATATACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1879
Db 1861 AGTGAATATATACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920
Oy 1880 CCTCATCTGATTTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1939
Db 1921 CCTCATCTGATTTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980
Oy 1940 AAAAAAGATTTTACTGAGAGAGAACTAATAAAGATTTAGTATGATGATGATGATG 1999
Db 1981 AAAAAAGATTTTACTGAGAGAGAACTAATAAAGATTTAGTATGATGATGATGATG 2040
Oy 2000 TTTGTAAGAGATGTTTGTCTACTGTCACCAATATGCTGCAAGTCTTCCAGCCCTGCT 2059
Db 2041 TTTGTAAGAGATGTTTGTCTACTGTCACCAATATGCTGCAAGTCTTCCAGCCCTGCT 2100
Oy 2060 GCTTCCAGAGACACAGAGAGAGAAATAGAGCATCCAGTATGATGATGATGATGATG 2119
Db 2101 GCTTCCAGAGACACAGAGAGAGAAATAGAGCATCCAGTATGATGATGATGATGATG 2160
Oy 2120 AGACATATGATGAG 2179
Db 2161 AGACATATGATGAG 2220
Oy 2180 AAATCTTCCAGATTTTCCAGATTTCCAGATTTCCAGATTTCCAGATTTCCAGATTT 2239
Db 2221 AAATCTTCCAGATTTTCCAGATTTCCAGATTTCCAGATTTCCAGATTTCCAGATTT 2280
Oy 2240 TGTAAATCTCAATATTAAG 2299
Db 2281 TGTAAATCTCAATATTAAG 2340
Oy 2300 ATTTACTCATGCTGTTAAATCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2359
Db 2341 ATTTACTCATGCTGTTAAATCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Oy 2360 ATATTTGATCTCTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2419
|||||

Db 2401 ATATTTGATCTCTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2460
Oy 2420 CCAAGTATTTGATCT 2479
Db 2461 CCAAGTATTTGATCT 2520
Oy 2480 GAAGAAATTAAG 2539
Db 2521 GAAGAAATTAAG 2580
Oy 2540 ACTCTCTCTTCCAGATTTATGATTTTCTAAGTATGATGATGATGATGATGATGATG 2599
Db 2581 ACTCTCTCTTCCAGATTTATGATTTTCTAAGTATGATGATGATGATGATGATGATG 2640
Oy 2600 CACTAGAGAGTTCCCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2659
Db 2641 CACTAGAGAGTTCCCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
Oy 2660 ACATGAGATTAATCCGAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2719
Db 2701 ACATGAGATTAATCCGAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Oy 2720 GAATGAGATTAATGCTTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2779
Db 2761 GAATGAGATTAATGCTTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Oy 2780 ATGCTCAAG 2839
Db 2821 ATGCTCAAG 2880
Oy 2840 GACTTAATTTATTTCTTAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2899
Db 2881 GACTTAATTTATTTCTTAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2940
Oy 2900 ATCCAAATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2959
Db 2941 ATCCAAATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
Oy 2960 AAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3019
Db 3001 AAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3060
Oy 3020 ACCCAAAATAG 3079
Db 3061 ACCCAAAATAG 3120
Oy 3080 AGCTTTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3139
Db 3121 AGCTTTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3180
Oy 3140 ATTTCAAGTTCATTTGCCAGATGATGATGATGATGATGATGATGATGATGATGATG 3199
Db 3181 ATTTCAAGTTCATTTGCCAGATGATGATGATGATGATGATGATGATGATGATGATG 3240
Oy 3200 GACAAAAATAG 3259
Db 3241 GACAAAAATAG 3300
Oy 3260 ACTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3319
Db 3301 ACTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3360
Oy 3320 TGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3379
Db 3361 TGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3420
Oy 3380 GAGACATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3439
Db 3421 GAGACATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3480
Oy 3440 ATCTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3499
Db 3481 ATCTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
|||||

OY	3500	TGCTGTGCTGCGCAAGCGCGAGTAACCTGTGTAAAGTAACCGTGGCAACAGTAGTT	3559
Db	3541	TGGTGTGCTGCGCAAGCGCGAGTAACCTGTGTAAAGTAACCGTGGCAACAGTAGTT	3600
OY	3560	TCCACAATTTCCCTGGTTCGCTCAGGTATCAAGTTGGAAAAATTTTGGCTATTAAAGAT	3619
Db	3601	TCCACAATTTCCCTGGTTCGCTCAGGTATCAAGTTGGAAAAATTTTGGCTATTAAAGAT	3660
OY	3620	ATGTATTAGATGCGTCTTATCCGATTTATTACTGGATACACTTGTATCTTTCTAATAT	3679
Db	3661	ATGTATTAGATGCGTCTTATCCGATTTATTACTGGATACACTTGTATCTTTCTAATAT	3720
OY	3680	TTCCAAAAGTGATGGGATTAACCTAGAAAGAGACACAGATATATTATTATTAAAGT	3739
Db	3721	TTCCAAAAGTGATGGGATTAACCTAGAAAGAGACACAGATATATTATTATTAAAGT	3780
OY	3740	AGCTTTAAAACCTCCTCTTATTATTCTACAAATTATATGCTAAATTTTCAGATTGAACAGG	3799
Db	3781	AGCTTTAAAACCTCCTCCTTATTCTTACAAAGTTATATGCTAAATTTTCAGATTGAACAGG	3840
OY	3800	ATTCCAGCATTCGCCATCCCATAGSAAAGAGAGGCTCCCATCTGAAAGGCTCCTGTA	3859
Db	3841	ATTCCAGCATTCGCCATCCCATAGSAAAGAGAGGCTCCCATCTGAAAGGCTCCTGTA	3900
OY	3860	AATCTACCCCTTCCAGAGCTTCAGACAAATCAGTTGATCTCCCGAGCGACACGCGCTCAT	3919
Db	3901	AATCTACCCCTTCCAGAGCTTCAGACAAATCAGTTGATCTCCCGAGCGACACGCGCTCAT	3960
OY	3920	CTGTGAGGAGGAGGAAAGATTAGCCAAAGAGTTAATTTTCATTCCAAATCAGTTAGCTGTT	3979
Db	3961	CTGTGAGGAGGAGGAAAGATTAGCCAAAGAGTTAATTTTCATTCCAAATCAGTTAGCTGTT	4020
OY	3980	AGACGATCTGTTTGTAGCAGTTGTTTGTATTTTGTCTGTGTCATTTTGTGAGACA	4039
Db	4021	AGACGATCTGTTTGTAGCAGTTGTTTGTATTTTGTCTGTGTCATTTTGTGAGACA	4080
OY	4040	TTTGTGTAGAAATTTCTATTTTGTGTCTACTGTATTTTCTTTTAAATATCTACTGAT	4099
Db	4081	TTTGTGTAGAAATTTCTATTTTGTGTCTACTGTATTTTCTTTTAAATATCTACTGAT	4140
OY	4100	ATCTTGTCCTTAAATTTTCTTCACATATAGTTTGCCTGATACCACTGATTTTATAACT	4159
Db	4141	ATCTTGTCCTTAAATTTTCTTCACATATAGTTTGCCTGATACCACTGATTTTATAACT	4200
OY	4160	GAATATTTAAGAAATCTAACAGCTAAACACTCAGTAAGTGATATTTTCTTATTAACATAG	4219
Db	4201	GAATATTTAAGAAATCTAACAGCTAAACACTCAGTAAGTGATATTTTCTTATTAACATAG	4260
OY	4220	ACCCGTTGCTACTCTCAGCACCCCTGCTCAATTTTTTCTGTAGCATGTGATGCTGTG	4279
Db	4261	ACCCGTTGCTACTCTCAGCACCCCTGCTCAATTTTTTCTGTAGCATGTGATGCTGTG	4320
OY	4280	ATTAAACATATTTCTATTGCTTTTATTTCATATATGGAACAATAGAGAGTAACTCTTA	4339
Db	4321	ATTAAACATATTTCTATTGCTTTTATTTCATATATGGAACAATAGAGAGTAACTCTTA	4380
OY	4340	ATATAGGTTGTAGTAAATTAACATCATTAAGCTAATATTATTAGAAAATGCTAATTAAGTAC	4399
Db	4381	ATATAGGTTGTAGTAAATTAACATCATTAAGCTAATATTATTAGAAAATGCTAATTAAGTAC	4440
OY	4400	CAGCACATAGAAAACATGAATTTGCTTAGTCATTTGACCTTTGTACGCAATTTTGAACAGTC	4459
Db	4441	CAGCACATAGAAAACATGAATTTGCTTAGTCATTTGACCTTTGTACGCAATTTTGAACAGTC	4500
OY	4460	ATTAAATGTTTGTCAATATTTTAAATAAAGTGCTGTGGTTTTCAGAAATACCTTC	4511
Db	4501	ATTAAATGTTTGTCAATATTTTAAATAAAGTGCTGTGGTTTTCAGAAATACCTTC	4552

DEFINITION	Homo sapiens cDNA FLJ12501 fis, clone NT2RM2001681.
ACCESSION	AK022563
VERSION	AK022563.1 GI:10434024
KEYWORDS	c11g0 capling; fis (full insert sequence).
SOURCE	Homo sapiens testicular carcinoma cell_line:NT2 CDNA to mRNA, clone_1lib:NT2RM2 clone:NT2RM2001681.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsumura,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahori,K., Masuno,Y. and Sasaki,N. NEDO human cDNA sequencing project Unpublished 2 (bases 1 to 2820) Isogai,T. and Otsuki,T. Direct Submission Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 297-0812, Japan (E-mail:genomich@helix.co.jp, Tel:01-438-52-3975, Fax:01-438-52-3986)
JOURNAL	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology; CDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES	Location/Qualifiers 1..2820 /organism="Homo sapiens" /db_xref="taxon:9606" /clone=NT2RM2001681 /cell_line="NT2" /cell_type="teratocarcinoma" /note="cloning vector: pME18SFU3-mRNA from uninduced NT2 neutonal precursor cells." 40..1113 /note="unnamed protein product" /codon_start=1 /protein_id="PAB14101.1" /db_xref="GI:10434025" translation="MYKKGEPVSYAGMLGTEDLKFTQLNRISYPVNITSQAEERY LSCGLKDLILYSSVSVALGESPMPKAKEDFSAGNTLKVITGVISEDDVLST KYAASLPALLARTBECKIESPLASHADIVQIITDALLENPEPTVENLPSFRL OKPLIFESDGTVPDYRKALITLVKKYIDLSFPCLWNLNKTPVGSLIKAYDPDLP APDLPLVINHSGGCVAFPPSDOAIIEENLVLMKTEIAGLENIHTTLPAOEKRPPE AYDLSEMDATDSORGRTRYKPKCKKEFDVQENDKEQHEDNSAVKKEPLETRITHMRNR SNMFKEAKSEFRDRKEJGCSKN"
CDS	BEST LOCAL SIMILARITY 99.8%; PRED. NO. 0; MISMATCHES 5; INDELS 0; GAPS 0;
BASE COUNT	911 a 496 c 504 g 909 t
ORIGIN	Query Match Best Local Similarity 99.8%; Pred. No. 0; Mismatched 5; Indels 0; Gaps 0;
Db	1 TGTACTAAGCAAAATGTTACTGAATTCCTTATCATTAAGAATGTACACAAGAGCGAGAAC 1745
Qy	1686 TGTACTAAGCAAAATGTTACTGAATTCCTTATCATTAAGAATGTACACAAGAGCGAGAAC 1745
Db	1 TGTACTAAGCAAAATGTTACTGAATTCCTTATCATTAAGAATGTACACAAGAGCGAGAAC 60
Qy	1746 CCAGTATTCTTGCTGGAATGTTAGAACAACAAGATCTCTAAAATTTATCCAGCTCAAC 1805
Db	61 CCAGTATTCTTGCTGGAATGTTAGAACAACAAGATCTCTAAAATTTATCCAGCTCAAC 120
Qy	1806 AGGATTTCATATCCAGTGAATATAATCAGATCCACAAGACAAGAATATTTAAGTGGG 1865
Db	121 AGGATTTCATATCCAGTGAATATAATCAGATCCACAAGACAAGAATATTTAAGTGGG 180
Qy	1866 GAATTTATTAANAAGCTCATCTTGATTTCTAGTGTGCATGATTAATGGGCTATTTAGTCCA 1925
Db	181 GAATTTATTAANAAGCTCATCTTGATTTCTAGTGTGCATGATTAATGGGCTATTTAGTCCA 240

OY	1926	ACCATGAAAAACAGCAAAAGAGATTTTAAgTGAAGCAGGAAACCTAATAAGATATGTT	1985
Db	241	ACCATGAAAAACAGCAAAAGAGATTTTAAgTGAAGCAGGAAACCTAATAAGATATGTT	300
OY	1986	ATCAGCTGGAATTTATTTCTGANAAGATGTTTTGCTACTGTCAACCAATATGCTGCAAGT	2045
Db	301	ATCAGCTGGAATTTATTTCTGANAAGATGTTTTGCTACTGTCAACCAATATGCTGCAAGT	360
OY	2046	CTTCCAGGCCCTGGTGGCTTGGCCACACACACGAAGGCGAAATATAGAGACCATCCCACTAGCT	2105
Db	361	CTTCCAGGCCCTGGTGGCTTGGCCACACACACGAAGGCGAAATATAGAGACCATCCCACTAGCT	420
OY	2106	AGCACACATGACACACACATAGTTCAATTAATTAACAGATGCAGTACGTAGAAATGTTTTCCG	2165
Db	421	AGCACACATGACACACACATAGTTCAATTAATTAACAGATGCAGTACGTAGAAATGTTTTCCG	480
OY	2166	GAAATCAGTGTGGAANAATCTTCCAGATTATTTGAGACTTCAGAAACCATTAATGATTTTG	2225
Db	481	GAAATCAGTGTGGAANAATCTTCCAGATTATTTGAGACTTCAGAAACCATTAATGATTTTG	540
OY	2226	TTTCAGAGATGCGACCTGTAATCTCTCAATATTAATAAAGCAATATTGACACGTGTAAAGCAG	2285
Db	541	TTTCAGAGATGCGACCTGTAATCTCTCAATATTAATAAAGCAATATTGACACGTGTAAAGCAG	600
OY	2286	AAATACCTTGATTCATTACTCTCCATCTGGTTAAATCTAAGAATATCTCCAGTGGGAGA	2345
Db	601	AAATACCTTGATTCATTACTCTCCATCTGGTTAAATCTAAGAATATCTCCAGTGGGAGA	660
OY	2346	GGAACTTGGGGCATATATTTTGATCTGTGCTGCCCTCCCTTCTCTTCTTGTGTTTTGGTGAAT	2405
Db	661	GGAACTTGGGGCATATATTTTGATCTGTGCTGCCCTCCCTTCTCTTCTTGTGTTTTGGTGAAT	720
OY	2406	CTGCATTCAGGTGGCCACATTTATTTGATTTCCCTGACAGCAGGCTAATAATGGAAGAAAC	2465
Db	721	CTGCATTCAGGTGGCCACATTTATTTGATTTCCCTGACAGCAGGCTAATAATGGAAGAAAC	780
OY	2466	CTTGTATTTGTGGCTGGAAGAAATTAGAAGCAGGACCTAGAAATCATATCACAAATTTTACCT	2525
Db	781	CTTGTATTTGTGGCTGGAAGAAATTAGAAGCAGGACCTAGAAATCATATCACAAATTTTACCT	840
OY	2526	GCTCAAGAAATGGAAACCTCTCTCTTCCAGCTTATGATTTTCTAAGATGATAGATGCCGA	2585
Db	841	GCTCAAGAAATGGAAACCTCTCTCTTCCAGCTTATGATTTTCTAAGATGATAGATGCCGA	900
OY	2586	ACATCTCAACGTGGCAGCTAGGAAGTTCCTCAAGTGTATGAAAAAACAGATGTGCAGAGAG	2645
Db	901	ACATCTCAACGTGGCAGCTAGGAAGTTCCTCAAGTGTATGAAAAAACAGATGTGCAGAGAG	960
OY	2646	AATGATTAAGGAACAACATGAAAGTAATGGGAGTGCAGAAAGAACCCGATTTGAAACCTGT	2705
Db	961	AATGATTAAGGAACAACATGAAAGTAATGGGAGTGCAGAAAGAACCCGATTTGAAACCTGT	1020
OY	2706	AGAATTAAGCATTTGGAATAGAAAGTAATGGTTTTAAAGAAAGCAGAAAAATCATTTAGACGT	2765
Db	1021	AGAATTAAGCATTTGGAATAGAAAGTAATGGTTTTAAAGAAAGCAGAAAAATCATTTAGACGT	1080
OY	2766	GATTAAGAAATTGAGATGCTCAAAAGGAACTATTTTATATGGGCTGTGGTTTTCCAAAATT	2825
Db	1081	GATTAAGAAATTGAGATGCTCAAAAGGAACTATTTTATATGGGCTGTGGTTTTCCAAAATT	1140
OY	2826	TTTTTGGCATGATAGACCTTAATTAATTTCTCTTAAGAAATTAATTAATCATTTCAAGTT	2885
Db	1141	TTTTTGGCATGATAGACCTTAATTAATTTCTCTTAAGAAATTAATTAATCATTTCAAGTT	1200
OY	2886	TGCAGACTAGTGCATCCAAATAGAAATTAATTAATTAAGTCCACATATTTTATTTAAATTTT	2945
Db	1201	TGCAGACTAGTGCATCCAAATAGAAATTAATTAATTAAGTCCACATATTTTATTTAAATTTT	1260
OY	2946	CTAGTACTACTATTTAAACAAAGTAAAGTGAAGCAGGCGCAAAATATATTTTCATATTTACTTT	3005
Db	1261	CTAGTACTACTATTTAAACAAAGTAAAGTGAAGCAGGCGCAAAATATATTTTCATATTTACTTT	1320

QY	3006	TCACCACTACTATATCCAAATAATGCGAAATATAGAAATATTTAATGACATATTTTACAT	3055
Db	1321	TCACCACTACTATATCCAAATAATGCGAAATATAGAAATATTTAATGACATATTTTACAT	1380
QY	3066	CCATTTTGTATCCAAAGCTCTTAATAGCAGTCATATTTTATACCTACGATTTTCTTAC	3125
Db	1381	CCATTTTGTATCCAAAGCTCTTAATAGCAGTCATATTTTATACCTACGATTTTCTTAC	1440
QY	3126	TTCCGAGTAGCCATATTTTCAAGGTTCATTTGCCACATGGCCCTGGACCTACTGTAATGG	3185
Db	1441	TTCCGAGTAGCCATATTTTCAAGGTTCCTTTGCCACATGGCCCTGTGACTACTGTAATGG	1500
QY	3186	ACAGTTAGCTACTAGACAAAACCTAGCATATTTAATTAAGTTTCTAGCATATTTCTATTT	3245
Db	1501	ACAGTTAGCTACTAGACAAAACCTAGCATATTTAATTAAGTTTCTAGCATATTTCTATTT	1560
QY	3246	TGGATTAAATTTAAACTCTAATCAGAGTTAATCTCCAGCTGATTCATATGCACTGACGT	3305
Db	1561	TGGATTAAATTTAAACTCTAATCAGAGTTAATCTCCAGCTGATTCATATGCACTGACGT	1620
QY	3306	TATATTTGTTTTATTTGGAGTCATGATATTTAAATTCAGGCTTTGTCAACCTCAAGGGAGAT	3365
Db	1621	TATATTTGTTTTATTTGGAGTCATGATATTTAAATTCAGGCTTTGTCAACCTCAAGGGAGAT	1680
QY	3366	TTAGCAATTTGTGCGAGACATTTTTTGTATGTATCATAGTACGAGTATTTAGCATTTAGTG	3425
Db	1681	TTAGCAATTTGTGCGAGACATTTTTTGTATGTATCATAGTACGAGTATTTAGCATTTAGTG	1740
QY	3426	AGTAGAGGCCATGATATCTCTGCTAAATATACCTGATTTGACACAGCCGCCACAACAAGAT	3485
Db	1741	AGTAGAGGCCACGGGATCTCTCTAAATATACCTGATTTGACACAGCCGCCACAACAAGAT	1800
QY	3486	TATCTGCGCCCAAAATGGTAGTCGCGCAAGGCGTGAATTAACCTTGTTTAAAAGTAACCTG	3545
Db	1801	TATCTGCGCCCAAAATGGTAGTCGCGCAAGGCGTGAATTAACCTTGTTTAAAAGTAACCTG	1860
QY	3546	TGGCAGACTAGGTTTCCAGAAATTTCTGCTTGTCTCATCGATATCATGTTTGA AAAAATTT	3605
Db	1861	TGGCAGACTAGGTTTCCAGAAATTTCTGCTTGTCTCATCGATATCATGTTTGA AAAAATTT	1920
QY	3606	TGGCATTTAAAGATATGATATTAAGTAGTGCTTATCTCGATATTTACCTGGATACAACTTGA	3665
Db	1921	TGGCATTTAAAGATATGATATTAAGTAGTGCTTATCTCGATATTTACCTGGATACAACTTGA	1980
QY	3666	TCTTTTCTAATATTTTACAGAAAGGATGAGGATTAACCTTGAGAGAGACATCAAGATGATAT	3725
Db	1981	TCTTTTCTAATATTTTACAGAAAGGATGAGGATTAACCTTGAGAGAGACTCAAGATGATAT	2040
QY	3726	TTATATTTTAAAGTAGAGCTTAAACCTCCCTCTATTTTCTACAAAGTTATATGGCTAAATTT	3785
Db	2041	TTATATTTTAAAGTAGAGCTTAAACCTCCCTCTATTTTCTACAAAGTTATATGGCTAAATTT	2100
QY	3786	CAGATTGAACAGGATTCAGCATTTCTGCAATCTCTCATGGAAGAAGAGCTCCCTCATTC	3845
Db	2101	CAGATTGAACAGGATTCAGCATTTCTGCAATCTCTCATGGAAGAAGAGCTCCCTCATTC	2160
QY	3846	TGAAGCGCTCTGAAATCTCAACCTTCGCAAGCTCAACAAACATCAGTGTATCCCTCGAC	3905
Db	2161	TGAAGCGCTCTGAAATCTCAACCTTCGCAAGCTCAACAAACATCAGTGTATCCCTCGAC	2220
QY	3906	CACACGGCTCATATCTGTGAGGAGGGAAGATTTAGCCCAAGAGTTAATTTTCATTTCCAA	3965
Db	2221	CACACGGCTCATATCTGTGAGGAGGGAAGATTTAGCCCAAGAGTTAATTTTCATTTCCAA	2280
QY	3966	ATCATTAGCTGTTAGACTATCTGTTTGTAGCAGTTGTTTGTCTCATATTTTGGCTGTGG	4025
Db	2281	ATCATTAGCTGTTAGACTATCTGTTTGTAGCAGTTGTTTGTCTCATATTTTGGCTGTGG	2340
QY	4026	CATTTTGTAGACATTTGTTGGAGATATTCATTTGGTGGCTCTACTGATATTTTCTTTT	4085
Db	2341	CATTTTGTAGACATTTGTTGGAGATATTCATTTGGTGGCTCTACTGATATTTTCTTTT	2400
QY	4086	AATATCTACTGATATCTTGTCTTTTAAATTTTCTTCAATATGCTTTGCCGATACAAAC	4145

Db	2401	AAATATCTACTTGGATATCTGTTCTTAAATTTCTTCACATFATGTTCCGTGATACAAAC	2460
OY	4146	TCATTTTAACTGAAATTTAAGCAATTAACAGCTAAACCTACGTAAAGTGCATMTATT	4205
Db	2461	TGATTTTATACTGAAATTTAAGCAATTAACAGCTAAACCTACGTAAAGTGCATMTATT	2520
OY	4206	TCCTTATACATAGAACCCGTTGCTACTGTCACACACCCTGCTCAATTTTTTTCCTGTA	4265
Db	2521	TGCTTATACATAGAACCCGTTGCTACTGTCACACACCCTGCTCAATTTTTTTCCTGTA	2580
OY	4266	GCATGTGATGCTGATTTAAACATCATTTTCATTTGCTTTTATTTTAAATVYGGGCAACATG	4325
Db	2581	GCATGTGATGCTGATTTAAACATCATTTTCATTTGCTTTTATTTTAAATVYGGGCAACATG	2640
OY	4326	AGAGGAACTCTAATATATAGTTGTAGTATATAAACATCATTTAGCTAATTTATAGAAAA	4385
Db	2641	AGAGGAACTCTAATATATAGTTGTAGTATATAAACATCATTTAGCTAATTTATAGAAAA	2700
OY	4386	TGCTTATATAGTACAGCAGCATAGAAACATGAAATTTGCTTACTCATTTGACTTTGTGCA	4445
Db	2701	TGCTTATATAGTACAGCAGCATAGAAACATGAAATTTGCTTACTCATTTGACTTTGTGCA	2760
OY	4446	CAATTTTGACATCATTTATTTTCTTCATTTTAAATTAATTAAGTGTGCTGCTTCAGAAAT	4505
Db	2761	CAATTTTGACATCATTTATTTTCTTCATTTTAAATTAATTAAGTGTGCTGCTTCAGAAAT	2820
RESULT 3			
AK097748			
LOCUS	AK097748	2747 bp	mRNA
DEFINITION	Homo sapiens CDNA FLJ040429 fls, clone TEST12039177.		linear
ACCESSION	AK097748		PRI 15-JUL-2002
VERSION	AK097748.1	GI:21577613	
KEYWORDS	Oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens testis CDNA to mRNA, clone_lib:TEST12		
ORGANISM	clone:TEST12039177.		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1		
AUTHORS	Osilima, A., Takahashi, Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,		
	Arta, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,		
	Otsuki, T., Sato, H., Wakematsu, A., Ishii, S., Yamamoto, J., Isono, Y.,		
	Kawai, Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,		
	Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,		
	Magatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,		
	Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and		
	Isogai, T.		
	NEO human cDNA sequencing project		
	unpublished		
	2 (bases 1 to 2747)		
	Isogai, T. and Yamamoto, J.		
	Direct Submission		
	Submitted (04-JUL-2002) Takao Isogai, FJ Project(HRI Team); 2-6-7		
	Kanusa Kamatari, Kisaizuru, Chiba 292-0812, Japan		
	(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)		
	NEO human cDNA sequencing project supported by Ministry of		
	Economy, Trade and Industry of Japan; cDNA full insert sequencing;		
	Research Association for Biotechnology (RAB); cDNA library		
	construction: Helix Research Institute (HRI) (supported by Japan		
	Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,		
	HRI, and Biotechnology Center, National Institute of Technology and		
	Evaluation; clone selection for full insert sequencing: HRI and		
	RAB; annotation: HRI and RAB.		
COMMENT			
JOURNAL			
AUTHORS			
TITLE			
FEATURES			
SOURCE			
	location/Qualifiers		
	1..2747		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="TEST12039177"		
	/issue_type="testis"		
	/clone_lib="TEST12"		
	/note="Cloning vector: pME18SF13"		

BASE COUNT	899 a	510 c	575 g	763 t	ORIGIN
Query Match	53.3%;	Score 2410.2;	DB 9;	Length 2747;	
Best Local Similarity	94.2%;	Pred. No. 0;			
Matches 2588;	Conservative	0;	Mismatches 3;	Indels 156;	Gaps 2
OY	139 GTTGGCCGGGGGGCCCTGGGGAACGGGCCCCGGGCCCCGGGAGGTCGGGGCAGATTAAC	198			
Db	1 GTTGGCCGGGGGGCCCTGGGGAACGGGCCCCGGGCCCCGGGAGGTCGGGGCAGATTAAC	60			
OY	199 ATAGATCATCAGTAGAAAACTCTTGAAGCTTGTCAAGAAAAATTTGAAAGTAGCAAAAT	258			
Db	61 ATAGATCATCAGTAGAAAACTCTTGAAGCTTGTCAAGAAAAATTTGAAAGTAGCAAAAT	120			
OY	259 AGAAAAATAAGAAATTTAACAGCAGATACAGAGACAGCATGGAAGTGTCTTAGCAAAC	318			
Db	121 AGAAAAATAAGAAATTTAACAGCAGATACAGAGACAGCATGGAAGTGTCTTAGCAAAC	180			
OY	319 AGAACACAGCAATGAAAAACACACAAAAATCCGGTCAGATACAACTGCGACGATATATG	378			
Db	181 AGAACACAGCAATGAAAAACACACAAAAATCCGGTCAGATACAACTGCGACGATATATG	240			
OY	379 TTTCCGGCTTCATGTCTTTAGAGTTGGGATCTCTTTGTCAATATGTCAATTTTACA	438			
Db	241 TTTCCGGCTTCATGTCTTTAGAGTTGGGATCTCTTTGTCAATATGTCAATTTTACA	300			
OY	439 TGCCAAACAGTAACCTCTTTACCAAGACGTGATCTCAGAAATATTTTACATTTGCAAC	498			
Db	301 TGCCAAACAGTAACCTCTTTACCAAGACGTGATCTCAGAAATATTTTACATTTGCAAC	360			
OY	499 CAGG-----TC	504			
Db	361 CAGGAAAAAGCCTCTTACGCTATATTTTGTCAAGCTGATTTCCCAAGAACATCTGATTTTC	420			
OY	505 TTGAGAACTGATGAGGCTGTTPAACACCTCTGCAGACACTGGAATTCAGTTGCCAAG	564			
Db	421 TTGAGAACTGATGAGGCTGTTPAACACCTCTGCAGACACTGGAATTCAGTTGCCAAG	480			
OY	565 TTTAATTTGTCTCAAGAGAAATATCAAGTACTGTGAAAAAGAAAAAGATTTGATGAAG	624			
Db	481 TTTAATTTGTCTCAAGAGAAATATCAAGTACTGTGAAAAAGAAAAAGATTTGATGAAG	540			
OY	625 CATATTTATTCAGAGGCAACATATTGCTCAGAGAAATCCCTACTGACACCTGTTGAG	684			
Db	541 CATATTTATTCAGAGGCAACATATTGCTCAGAGAAATCCCTACTGACACCTGTTGAG	600			
OY	685 TGAATGCCATGTGCCCCATGTTCTTGTGCTCTCTTTTGTGTAAGGAATATATTA	744			
Db	601 TGAATGCCATGTGCCCCATGTTCTTGTGCTCTCTTTTGTGTAAGGAATATATTA	660			
OY	745 CCAACCTGGAGACCTTGAGAACATAGAAAAATGCTGTGAAGAAAAAGCAATATATAT	804			
Db	661 CCAACCTGGAGACCTTGAGAACATAGAAAAATGCTGTGAAGAAAAAGCAATATATAT	720			
OY	805 TCTCATATGTAAAGACCATTTGGAATACAGACACAGACAGTCAATGGAACCGGTTTG	864			
Db	721 TCTCATATGTAAAGACCATTTGGAATACAGACACAGACAGTCAATGGAACCGGTTTG	780			
OY	865 TGTATGGAGACTACATACCAATTTGTCTTAACCAAGAAATGGCCCTTTTGGAAACATATG	924			
Db	781 TGTATGGAGACTACATACCAATTTGTCTTAACCAAGAAATGGCCCTTTTGGAAACATATG	840			
OY	925 GCTCTGAGAGTGTGAATATGACATCTCTACTTTTTCATATGTAAACAGTCTTGACT	984			
Db	841 GCTCTGAGAGTGTGAATATGATATCTCTACTTTTTCATATGTAAACAGTCTTGACT	900			
OY	985 TGACCCAGCAATGTAGAGAAACACTAATGTGAACAGCCATTGCATCACTGAACATTCACC	1044			
Db	901 TGACCCAGCAATGTAGAGAAACACTAATGTGAACAGCCATTGCATCACTGAACATTCACC	960			
OY	1045 TGTATTATTAAGACATGGAAGCAGCTCTGTGTGACTGGAAGTGTCTGAACATCTCTCAACAG	1104			

Db 961 TGTATTATTAAGACAATGAAGACACCTCTGTGACTGAATGCTGAGATCTCAACAAG 1020
 Qy 1105 TTTCACCTCTCCATCCCAACCTGGGCTTACCACTGGTTTATTTGTATACCAACAAGGCTA 1164
 Db 1021 TTTCACCTCTCCATCCCAACCTGGGCTTACCACTGGTTTATTTGTATACCAACAAGGCTA 1080
 Qy 1165 CTTATGAAGCTATGAAAGAACTGAGAAATGGGTTGCTTGGGCTCTTGGGAAAAAGCAG 1224
 Db 1081 CTTATGAAGCTATGAAAGAACTGAGAAATGGGTTGCTTGGGCTCTTGGGAAAAAGCAG 1140
 Qy 1235 GAGTTCTACTCTTTTAAAGGAGCTCTTGGAGTGAACATCTTCAAGATCTAATGTGG 1284
 Db 1141 GAGTTCTACTCTTTTAAAGGAGCTCTTGGAGTGAACATCTTCAAGATCTAATGTGG 1200
 Qy 1285 TCTTCAAAAAGACAGAAAGGAGGAGTTCCAGTGAATTTTGGTATTTACATGATGTGAT 1344
 Db 1201 TCTTCAAAAAGACAGAAAGGAGGAGTTCCAGTGAATTTTGGTATTTACATGATGTGAT 1260
 Qy 1345 TAATTAATATCTCATGTGGAATAATATGACATTGAGAAATACAGAAGATGAAGACA 1404
 Db 1261 TAATTAATATCTCATGTGGAATAATATGACATTGAGAAATACAGAAGATGAAGACA 1320
 Qy 1405 ATGACATGGAAGTCCAGATATAGATGTTCAAGATGATGAATGGCAGAACTGTTTCA 1464
 Db 1321 ATGACATGGAAGTCCAGATATAGATGTTCAAGATGATGAATGGCAGAACTGTTTCA 1380
 Qy 1465 GAGATGAGGAAGAAATTAATCTTGGAACTAGACATGAGTGAACAGAAAGAACTTTA 1524
 Db 1381 GAGATGAGGAAGAAATTAATCTTGGAACTAGACATGAGTGAACAGAAAGAACTTTA 1440
 Qy 1525 ATGCAACAGTATGAGCTTCTGACAGCATTAGTACTCTTATGCTGGTGGCAAGAGAT 1584
 Db 1441 ATGCAACAGTATGAGCTTCTGACAGCATTAGTACTCTTATGCTGGTGGCAAGAGAT 1500
 Qy 1585 CCATGGCATTTTTCGAATCTTATATTGATGTGGCAGTTAACTGAAA----- 1631
 Db 1501 CCATGGCATTTTTCGAATCTTATATTGATGTGGCAGTTAACTGAAAAGTTGCTGGGCGC 1560
 Qy 1632 ----- 1631
 Db 1561 ATCTTGTCTCTCTTACACGACACATGGCTGGCAGAGAGGAAAGACAGAGCGGCATC 1620
 Qy 1632 -----GGCAGATCTACTATGCTTCTTCTTACTAGAAATA 1662
 Db 1621 TTGGAACGTCATGTCCTTACTTCTGCGGCGACATCTACTATGCTTCTTACTAGAAATA 1680
 Qy 1663 ACTGTCAGATGTCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1722
 Db 1681 ACTGTCAGATGTCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 Qy 1723 AGATGTACAGAAGAGCGAGAACCCAGTATCTTATGCTGGATGTAGGAACCAAGAATC 1782
 Db 1741 AGATGTACAGAAGAGCGAGAACCCAGTATCTTATGCTGGATGTAGGAACCAAGAATC 1800
 Qy 1783 TCTTAAATTTATCCAGCTCAACAGATTTCAATCCAGTGAATTAACATGATCCAG 1842
 Db 1801 TCTTAAATTTATCCAGCTCAACAGATTTCAATCCAGTGAATTAACATGATCCAG 1860
 Qy 1843 AAGCAGAAGATTTAAGTGGGAATTAATAAAGACTCATCTTGTATTCTAGTGTGT 1902
 Db 1861 AAGCAGAAGATTTAAGTGGGAATTAATAAAGACTCATCTTGTATTCTAGTGTGT 1920
 Qy 1903 CAGTATGAGGACTATTTGTCACACCAAGTAAGAAAGCAAAAGATTTTACTGAAGCAG 1962
 Db 1921 CAGTATGAGGACTATTTGTCACACCAAGTAAGAAAGCAAAAGATTTTACTGAAGCAG 1980
 Qy 1963 GAACTACTTAAAGATATGTTATCAGTGAATTTATCTGAAGAAGATGTTTGTCTAC 2022
 Db 1981 GAACTACTTAAAGATATGTTATCAGTGAATTTATCTGAAGAAGATGTTTGTCTAC 2040
 Qy 2023 TGTCAACAATAATATGCTGCAAGCTTCCAGCCCTGCTGCTCCAGACACACAGAAAGCA 2082
 Db 2041 TGTCAACAATAATATGCTGCAAGCTTCCAGCCCTGCTGCTCCAGACACACAGAAAGCA 2100

Qy 2083 AATAGAGAGCATCCCTAGCTAGCAGACATGCAACAGACATAGTCAATATAACAG 2142
 Db 2101 AATAGAGAGCATCCCTAGCTAGCAGACATGCAACAGACATAGTCAATATAACAG 2160
 Qy 2143 ATGCACTACTGGAATAATGTTCCGAAATCATCTGGAATAATCTCCAGTATTTTCAGAC 2202
 Db 2161 ATGCACTACTGGAATAATGTTCCGAAATCATCTGGAATAATCTCCAGTATTTTCAGAC 2220
 Qy 2203 TTGAGAACCATTATGATTTTGTTCAGTATGCGACCTGTAAATCCTCAATATAAAG 2262
 Db 2221 TTGAGAACCATTATGATTTTGTTCAGTATGCGACCTGTAAATCCTCAATATAAAG 2280
 Qy 2263 CATATTTGACACTGCTGAAGCAGAAATTAATGATTTCAATTTCTCATGCTGTTAAAC 2322
 Db 2281 CATATTTGACACTGCTGAAGCAGAAATTAATGATTTCAATTTCTCATGCTGTTAAAC 2340
 Qy 2323 TAAAGATACCTCAGTGGGAGAGGAATCTTGGGGCATATTTGATCCTGCTGCC 2382
 Db 2341 TAAAGATACCTCAGTGGGAGAGGAATCTTGGGGCATATTTGATCCTGCTGCC 2400
 Qy 2383 TTCTCTTCTTGTGTTGGTGAATCTGCAATTCAGTGTGGCCAAATTTTGCATTTCTCAG 2442
 Db 2401 TTCTCTTCTTGTGTTGGTGAATCTGCAATTCAGTGTGGCCAAATTTTGCATTTCTCAG 2460
 Qy 2443 ACCAGGCTATTAATGGAAGAAACCTGTAATGCTGGAAGAAATGGAAGCAGACTAG 2502
 Db 2461 ACCAGGCTATTAATGGAAGAAACCTGTAATGCTGGAAGAAATGGAAGCAGACTAG 2520
 Qy 2503 AAAATCATATCACAATTTTACCTGCTCAAGAAATGGAAGAACTCCTCTTCCAGCTTATGAT 2562
 Db 2521 AAAATCATATCACAATTTTACCTGCTCAAGAAATGGAAGAACTCCTCTTCCAGCTTATGAT 2580
 Qy 2563 TTCTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2622
 Db 2581 TTCTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
 Qy 2623 TGAAGAAGACAGATGTCAGAGGAATGATGAGAACATGATGATGATGATGATGATGATGAT 2682
 Db 2641 TGAAGAAGACAGATGTCAGAGGAATGATGAGAACATGATGATGATGATGATGATGATGAT 2700
 Qy 2683 GAAAGAAGACGATGTAACCTGAGAAATTAAGCATTTGAATGAAGT 2729
 Db 2701 GAAAGAAGACGATGTAACCTGAGAAATTAAGCATTTGAATGAAGT 2747

RESULT 4
 AC025659/C
 LOCUS
 DEFINITION
 Homo sapiens clone RP11-263N17, WORKING DRAFT SEQUENCE, 15
 unorderd pieces.
 AC025659
 VERSION
 AC025659.3 GI:8705128
 KEYWORDS
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 Homo sapiens.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 161100)
 Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-263N17
 JOURNAL
 Unpublished
 2 (bases 1 to 161100)
 Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
 Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Plerre,N., Grant,G., Hago,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kamp,L., Karatas,A.,
 Klein,J., Labroque,K., Lamazares,R., Landers,T., Lehoczy,J.,

TITLE
JOURNAL
COMMENT

Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPhee, R., Meldrum, J., Meneu, L., Mihova, T., Mihaeva, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Ollivar, T. M., Oliver, J., Peterson, K., Pierre, N., Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced g1:7651991.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7940

Center clone name: 263_N_17

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 153679 bases at least Q40

Consensus quality: 157422 bases at least Q30

Consensus quality: 158798 bases at least Q20

Insert size: 16300; agarose-fp

Insert size: 159700; sum-of-ctrls

Quality coverage: 5.1 in Q20 bases; agarose-fp

Quality coverage: 5.2 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      3393: contig of 3393 bp in length
*      3394 3493: gap of 100 bp
*      3494 8261: contig of 4768 bp in length
*      8262 8361: gap of 100 bp
*      8362 13937: contig of 5576 bp in length
*      13938 14037: gap of 100 bp
*      14038 23507: contig of 9470 bp in length
*      23508 23607: gap of 100 bp
*      23608 30363: contig of 6756 bp in length
*      30364 30463: gap of 100 bp
*      30464 39080: contig of 8617 bp in length
*      39081 39180: gap of 100 bp
*      39181 47753: contig of 8573 bp in length
*      47754 47853: gap of 100 bp
*      47854 59271: contig of 11418 bp in length
*      59272 59371: gap of 100 bp
*      59372 71027: contig of 11656 bp in length
*      71028 71127: gap of 100 bp
*      71128 84959: contig of 13832 bp in length
*      84960 85059: gap of 100 bp
*      85060 95488: contig of 10429 bp in length
*      95489 95588: gap of 100 bp
*      95589 108075: contig of 12487 bp in length
*      108076 108175: gap of 100 bp
*      108176 123260: contig of 15085 bp in length
*      123261 123360: gap of 100 bp
*      123361 140393: contig of 17033 bp in length

```

```

FEATURES             * 140394 140493: gap of 100 bp
                      * 140494 161100: contig of 20607 bp in length.
                      Location/Qualifiers
                        source
                          1. 161100
                             /organism="Homo sapiens"
                             /db_xref="taxon:9606"
                             /clone="RP11-263N1"
                             /clone_lib="RP11 Human Male BAC"
                             1. 3393
                                /note="assembly-fragment"
                                3494..8261
                                   /note="assembly-fragment"
                                   8362..13937
                                      /note="assembly-fragment"
                                      14038..23507
                                         /note="assembly-fragment"
                                         23608..30363
                                            /note="assembly-fragment"
                                            vector_side:right"
                                              clone_end:Sp6
                                                 30464..39080
                                                    /note="assembly-fragment"
                                                    39181..47753
                                                       /note="assembly-fragment"
                                                       47854..59271
                                                          /note="assembly-fragment"
                                                          59372..71027
                                                             /note="assembly-fragment"
                                                             71128..84959
                                                                /note="assembly-fragment"
                                                                85060..95488
                                                                   /note="assembly-fragment"
                                                                   95589..108075
                                                                      /note="assembly-fragment"
                                                                      108176..123260
                                                                         /note="assembly-fragment"
                                                                         123361..140393
                                                                            /note="assembly-fragment"
                                                                            140494..161100
                                                                               /note="assembly-fragment"

```

```

BASE COUNT  47798 a 31689 c 31601 g 48609 t 1403 others
ORIGIN

```

```

Query Match          44.1%; Score 1995.8; DB 2; Length 161100;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1997; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2515 CAATTTACCTGCTCAAGATGGAACCTCCTCCAGCTTATGATTTCTAAGTATGA 2574
      |||||||
Db 114956 CAATTTACCTGCTCAAGATGGAACCTCCTCCAGCTTATGATTTCTAAGTATGA 114897

QY 2575 TAGATGCCGCAACATCTCAACGTCGACATGGAAGTTCCCAAGTGTATGAAGAAACAG 2634
      |||||||
Db 114896 TAGATGCCGCAACATCTCAACGTCGACATGGAAGTTCCCAAGTGTATGAAGAAACAG 114837

QY 2635 ATGTGCAGAGAAATGATTAAGCAACACATGATTAATGCGAGTCAGCAAGAACCGCA 2694
      |||||||
Db 114836 ATGTGCAGAGAAATGATTAAGCAACACATGATTAATGCGAGTCAGCAAGAACCGCA 114777

QY 2695 TTGAACCTGAGCAATTAACATTTGATGAGTATGTTTAAAGACGAAATAAT 2754
      |||||||
Db 114776 TTGAACCTGAGCAATTAACATTTGATGAGTATGTTTAAAGACGAAATAAT 114717

QY 2755 CATTTAGACGTGATTAAGATTAAGATGATGCTCAAAAGTGAACATTAATTTATAGGCGCTGCG 2814
      |||||||
Db 114716 CATTTAGACGTGATTAAGATTAAGATGATGCTCAAAAGTGAACATTAATTTATAGGCGCTGCG 114657

QY 2815 TTTCACAAATTTTTTGGCATGATAGACTTAATTTATTTCTTAAGCAATATTAAT 2874
      |||||||
Db 114656 TTTCACAAATTTTTTGGCATGATAGACTTAATTTATTTCTTAAGCAATATTAAT 114597

QY 2875 CATTTCAGTTTGACAGCTAGTGCACATCCCAATGAATTAATTAATAGTCACATATTTTA 2934
      |||||||

```


Db 114596 CATTCAGTTTGCAGACTAGTCCATCCATAGAAATTAATATAAGTCACATATTTTA 114537
 QY 2935 TTTAAATTTTCTAGTACTACATTAACAAGATAAAGGAGGAGGCAAAATTTT 2994
 Db 114536 TTTAAATTTTCTAGTACTACATTAACAAGATAAAGGAGGAGGCAAAATTTT 114477
 QY 2995 GATATTACTTTTCCACCAGTAGATATACCAAAATAGCGAAATATAGAAATTTAATGAG 3054
 Db 114476 GATATTACTTTTCCACCAGTAGATATACCAAAATAGCGAAATATAGAAATTTAATGAG 114417
 QY 3055 AATTTTAACTGCTTTTGTACCAAGTCTTAAATGCAAGTACATATTTTACTTACT 3114
 Db 114416 AATTTTAACTGCTTTTGTACCAAGTCTTAAATGCAAGTACATATTTTACTTACT 114357
 QY 3115 GCATTTCTACTTCCAGTAGCCATATTTCAAGTGTTCATTTGCCACATGGGCTGTGAC 3174
 Db 114356 GCATTTCTACTTCCAGTAGCCATATTTCAAGTGTTCATTTGCCACATGGGCTGTGAC 114297
 QY 3175 TACTGTATTGAGAGTTCAGTACTAGACAAAACATAGCATTAATTAATTAGTTCTAGCCA 3234
 Db 114296 TACTGTATTGAGAGTTCAGTACTAGACAAAACATAGCATTAATTAATTAGTTCTAGCCA 114237
 QY 3235 TGAATTTCTATTGGATTTAAATTAACCTTAATCAGAGTTAACTCCACAGTGCATTCATG 3294
 Db 114236 TGAATTTCTATTGGATTTAAATTAACCTTAATCAGAGTTAACTCCACAGTGCATTCATG 114177
 QY 3295 CAGCTCAGAGTTATATTTTGTATTTGAGTCTGATATTAATAACAGGTTTGTCAACC 3354
 Db 114176 CAGCTCAGAGTTATATTTTGTATTTGAGTCTGATATTAATAACAGGTTTGTCAACC 114117
 QY 3355 TCAGGGGATATTATAGCAATTTGTGGGAGACATTTTGTATGTCAATGACAGGAGTATTT 3414
 Db 114116 TCAGGGGATATTATAGCAATTTGTGGGAGACATTTTGTATGTCAATGACAGGAGTATTT 114057
 QY 3415 GACATTTAGTGTAGAGGCCATGATCCTGCTAAATTAACCTGATTTGGACAGGCCCA 3474
 Db 114056 GACATTTAGTGTAGAGGCCATGATCCTGCTAAATTAACCTGATTTGGACAGGCCCA 113997
 QY 3475 CACAAAGAAATATATCTGCGCCGAATGTAGTGTGCCAAGGCTGATTAACCTTGTGTTA 3534
 Db 113996 CACAAAGAAATATATCTGCGCCGAATGTAGTGTGCCAAGGCTGATTAACCTTGTGTTA 113937
 QY 3535 AAAGTAACTGTGGCAGACTAGGTTTCCAGAAATTTCTGGTTCTGCTCAAGTATCTGTT 3594
 Db 113936 AAAGTAACTGTGGCAGACTAGGTTTCCAGAAATTTCTGGTTCTGCTCAAGTATCTGTT 113877
 QY 3595 TGAATAAATTTTGGCTATTAAAGATATGTATTAGATGTCTTATCTGATTTTACCTGG 3654
 Db 113876 TGAATAAATTTTGGCTATTAAAGATATGTATTAGATGTCTTATCTGATTTTACCTGG 113817
 QY 3655 ATACAACTGATCTTTCTAATATTTTCAAGAAAGTATGGGATTAACCTTAAGAAAGAGACT 3714
 Db 113816 ATACAACTGATCTTTCTAATATTTTCAAGAAAGTATGGGATTAACCTTAAGAAAGAGACT 113757
 QY 3715 CAGAATGATATTATTTTAAAGTGTAAACCTCTCTTATTTCTACAAAGTTATA 3774
 Db 113756 CAGAATGATATTATTTTAAAGTGTAAACCTCTCTTATTTCTACAAAGTTATA 113697
 QY 3775 TGGCTAATTTTTCAGATTGAGAGGATTCAGATTCCTGATTCCTCATGGAAGAGAG 3834
 Db 113696 TGGCTAATTTTTCAGATTGAGAGGATTCAGATTCCTGATTCCTCATGGAAGAGAG 113637
 QY 3835 GGTCCCTCATCTGAAGCGTCTGTGAATATTAACCTTGCAGAGCTTCAGACAAATCACTTGA 3894
 Db 113636 GGTCCCTCATCTGAAGCGTCTGTGAATATTAACCTTGCAGAGCTTCAGACAAATCACTTGA 113577
 QY 3895 TCTCCCTGAGCCACAGCGGCTCATTTCTGAGAGGAGAAAGATTAGCCAAAGAGTTAAT 3954
 Db 113576 TCTCCCTGAGCCACAGCGGCTCATTTCTGAGAGGAGAAAGATTAGCCAAAGAGTTAAT 113517
 QY 3955 TTTTCATTTCCAAATCACTTATGCTGTAGACTGATCTGTGTTGTAGCAGTTGTTGTCTCAT 4014
 Db 113516 TTTTCATTTCCAAATCACTTATGCTGTAGACTGATCTGTGTTGTAGCAGTTGTTGTCTCAT 113457

QY 4015 TTTGCTGTGCAATTTTGTGAGACATTTGTGGAATATTTCTAATTTGGTGTCTACTGTA 4074
 Db 113456 TTTGCTGTGCAATTTTGTGAGACATTTGTGGAATATTTCTAATTTGGTGTCTACTGTA 113397
 QY 4075 TTTTCTTTTAAATCTACTTATATCTTGTCTTTAAATTTTCTTACATATGTTTG 4134
 Db 113396 TTTTCTTTTAAATCTACTTATATCTTGTCTTTAAATTTTCTTACATATGTTTG 113337
 QY 4135 CCTGATACAACTGATTTTATTAATCTGAATTTTAAAGATTTTACAGCTAAACTCAGTAA 4194
 Db 113336 CCTGATACAACTGATTTTATTAATCTGAATTTTAAAGATTTTACAGCTAAACTCAGTAA 113277
 QY 4195 GTGCATATATTTCTTATTAATCAGACCCGCTCTACTCTCAGACCCCTCTCTCAATTT 4254
 Db 113276 GTGCATATATTTCTTATTAATCAGACCCGCTCTACTCTCAGACCCCTCTCTCAATTT 113217
 QY 4255 TTTTTCCTGTAGCATGTATGCTGTGAATTAACATTTTCAATTTGCTTTTATTTCTAATA 4314
 Db 113216 TTTTTCCTGTAGCATGTATGCTGTGAATTAACATTTTCAATTTGCTTTTATTTCTAATA 113157
 QY 4315 TGGGAACAAATGAGAGTGAATCTTAATATATGTTGTAGTATTAATAACATCATAGCCTAA 4374
 Db 113156 TGGGAACAAATGAGAGTGAATCTTAATATATGTTGTAGTATTAATAACATCATAGCCTAA 113097
 QY 4375 TTTTGAATAATGCTAATTAATGATACAGACATAGAAACATGAATTTGCTTACTATTTGT 4434
 Db 113096 TTTTGAATAATGCTAATTAATGATACAGACATAGAAACATGAATTTGCTTACTATTTGT 113037
 QY 4435 ACCTTTGTGACAAATTTTGCAGCTCATTAATGTTTGTCTATTAATTTAATAAGTGTCTG 4494
 Db 113036 ACCTTTGTGACAAATTTTGCAGCTCATTAATGTTTGTCTATTAATTTAATAAGTGTCTG 112977
 QY 4495 GGTTCAGAAATACCTTCAAAA 4515
 Db 112976 GGTTCAGAAATACCTTCAAAA 112956

RESULT 5
 CDS1RGC
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 176368)
 Heilig, R., Petlic, J.L., Vico, V., Desilva, C., Robert, C., Wnicker, P.,
 Brothier, P., Catolico, L., Barde, V., Pellelier, E., Artiguenave, F.,
 Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
 Gyapay, G., Saurin, W. and Weissbach, J.
 JOURNAL
 TITLE
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 176368)
 Genoscope.
 Direct Submission
 Submitted (08-NOV-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
 On Nov 15, 2001 this sequence version replaced gi:14715170.
 - Web : www.genoscope.cns.fr)
 Center: Genoscope / Centre National de Sequencage
 Center: Genoscope
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: Segref@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
 ----- Summary Statistics
 Assembly program: Phrap; version 2.0

Quality coverage: 7.53x in 020 bases; sum-of-ctrls

Overall quality chart :
Range : bases
0 - 9 : 5
10 - 19 : 25
20 - 29 : 67
30 - 39 : 185
40 - 49 : 2535
50 - 59 : 5835
60 - 69 : 7981
70 - 79 : 20287
80 - 89 : 49616
90 - 99 : 89832

Percentage of bases with a quality value >= 40 : 99 %

FEATURES
Source Location/Qualifiers
1. 176368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-93311"
/clone_lib="RBC1-11"
BASE COUNT 49878 a 31876 c 34551 g 60063 t
ORIGIN
Query Match 44.1%; Score 1995.8; DB 9; Length 176368;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1997; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2515 CAATTTACCTGCTCAAGAAATGGAACCTCTTCAGCTTATGATTTCTTAAGTATGA 2574
Db 154334 CAATTTACCTGCTCAAGAAATGGAACCTCTTCAGCTTATGATTTCTTAAGTATGA 154393
QY 2575 TAGATGCGGCAACATCAACGCTGACCTAGGAAGTTCCCAAGTATGAAAGAAACAG 2634
Db 154394 TAGATGCGGCAACATCAACGCTGACCTAGGAAGTTCCCAAGTATGAAAGAAACAG 154453
QY 2635 ATGTGCGAGAAATGATAGGAACACATGATGAATTAATGCGCAGTCAGAAAGAACCGA 2694
Db 154454 ATGTGCGAGAAATGATAGGAACACATGATGAATTAATGCGCAGTCAGAAAGAACCGA 154513
QY 2695 TTGAACCTGAGAAATGAACGATTGGAATGAAGTAATGGTTTAAAGAACAGAAAAAT 2754
Db 154514 TTGAACCTGAGAAATGAACGATTGGAATGAAGTAATGGTTTAAAGAACAGAAAAAT 154573
QY 2755 CATTTAGCGTGAATGAAGTATGAGTCAAAAGTGAAGTATTTTATAGGCTGTGG 2814
Db 154574 CATTTAGCGTGAATGAAGTATGAGTCAAAAGTGAAGTATTTTATAGGCTGTGG 154633
QY 2815 TTTCACAAATTTTGGTGGATGATGACTTAATTTATTCCTTAAAGAAATATTTAAAT 2874
Db 154634 TTTCACAAATTTTGGTGGATGATGACTTAATTTATTCCTTAAAGAAATATTTAAAT 154693
QY 2875 CATTTCAAGTTTGAGACTAGTGCATTCATAGAAATTAATATTAAGTACATATTTTA 2934
Db 154694 CATTTCAAGTTTGAGACTAGTGCATTCATAGAAATTAATATTAAGTACATATTTTA 154753
QY 2935 TTTTAAATTTTCTAGTAACTAATTAACAAAGTAAGGAGGAGGCAAAATATTTT 2994
Db 154754 TTTTAAATTTTCTAGTAACTAATTAACAAAGTAAGGAGGAGGCAAAATATTTT 154813
QY 2995 GATATTACTTTTCAACAGTATATACCAAAATAGGCAATATAGAAATTTATTAATGAG 3054
Db 154814 GATATTACTTTTCAACAGTATATACCAAAATAGGCAATATAGAAATTTATTAATGAG 154873
QY 3055 ATATTATACCTCTTTTGTACCAAGCTCTTAATAGCAGTACATATTTTATACCTACT 3114
Db 154874 ATATTATACCTCTTTTGTACCAAGCTCTTAATAGCAGTACATATTTTATACCTACT 154933
QY 3115 GCATTTCTTACTTCGAGTAGCACAATTTTCAAGTGTTCATGCCACATGTGGCCTGTGAC 3174

|||||
Db 154934 GCATTTCTTACTTCGAGTAGCACAATTTTCAAGTGTTCATGCCACATGTGGCCTGTGAC 154993
QY 3175 TACTGATTTGAGACAGTTCAGTACTAGACAAACATACATATTTAACTAGTTCTAGCCA 3234
Db 154994 TACTGATTTGAGACAGTTCAGTACTAGACAAACATACATATTTAACTAGTTCTAGCCA 155053
QY 3235 TGATTTCTATTTGGATTTAAATTAACCTATATACAGATTAACTCCACAGTGCATTCATG 3294
Db 155054 TGATTTCTATTTGGATTTAAATTAACCTATATACAGATTAACTCCACAGTGCATTCATG 155113
QY 3295 CAGCTGACAGTATATTTGTTTATTTGAGTGCATGATATTTAAATCAGCCTTTGTCAACC 3354
Db 155114 CAGCTGACAGTATATTTGTTTATTTGAGTGCATGATATTTAAATCAGCCTTTGTCAACC 155173
QY 3355 TCAGGGATATTTAGCAATTTGTGCGAGACATTTTGTGATTCATGACTAGGCAAGTTATT 3414
Db 155174 TCAGGGATATTTAGCAATTTGTGCGAGACATTTTGTGATTCATGACTAGGCAAGTTATT 155233
QY 3415 GACATTTAGTAGTAGAGGCAATGATCTGTAAATTAACCTGCATTTGACAGCGCCCA 3474
Db 155234 GACATTTAGTAGTAGAGGCAATGATCTGTAAATTAACCTGCATTTGACAGCGCCCA 155293
QY 3475 CAACAAAGATTTATCCCTGCCGAAATGTGTGTGTGCGCAAGGCTGAGTAACCTTGTGTTA 3534
Db 155294 CAACAAAGATTTATCCCTGCCGAAATGTGTGTGTGCGCAAGGCTGAGTAACCTTGTGTTA 155353
QY 3535 AAAGTAACCTGTGCGACAGTATGTTTCCAGAAATTTCCGTGTTGCTGACAGTATCATGTT 3594
Db 155354 AAAGTAACCTGTGCGACAGTATGTTTCCAGAAATTTCCGTGTTGCTGACAGTATCATGTT 155413
QY 3595 TGAATAAATTTTGGCTATTTAAAGATATGATTTAGATGCTTTATCCGATTTATACCTGG 3654
Db 155414 TGAATAAATTTTGGCTATTTAAAGATATGATTTAGATGCTTTATCCGATTTATACCTGG 155473
QY 3655 ATCAACTGATCTTTTCTATATATTTTCAAGAAATGAGTGAATACCTTGAAGAGACT 3714
Db 155474 ATCAACTGATCTTTTCTATATATTTTCAAGAAATGAGTGAATACCTTGAAGAGACT 155533
QY 3715 CAGAAATATTTATATTTTAAAGTACTTAAACCTCTCTTATTTCTACAAGTTATA 3774
Db 155534 CAGAAATATTTATATTTTAAAGTACTTAAACCTCTCTTATTTCTACAAGTTATA 155593
QY 3775 TGCGTAAATTTCAAGTTGAACAGGATTCAGCATTCGCCATCTCCCTCATGAGAAAGAG 3834
Db 155594 TGCGTAAATTTCAAGTTGAACAGGATTCAGCATTCGCCATCTCCCTCATGAGAAAGAG 155653
QY 3835 GCTCCCTCATCTGAGGCGTCTGGAATCTACCTTGCAGAGTTCAGACAAATCAGTTGA 3894
Db 155654 GCTCCCTCATCTGAGGCGTCTGGAATCTACCTTGCAGAGTTCAGACAAATCAGTTGA 155713
QY 3895 TCTCCCTGAGCCACAGCGCTCATTTCTGTGAGGAGGAGGAAAGATTAGCCAAAGATTAAAT 3954
Db 155714 TCTCCCTGAGCCACAGCGCTCATTTCTGTGAGGAGGAGGAAAGATTAGCCAAAGATTAAAT 155773
QY 3955 TTTCATTTCCAAATCAGTACTAGCTGTAGACTGATCTGTTTGTACAGTTGTTGTCTCAT 4014
Db 155774 TTTCATTTCCAAATCAGTACTAGCTGTAGACTGATCTGTTTGTACAGTTGTTGTCTCAT 155833
QY 4015 TTTGCTGTGCAATTTTGTGAGACATTTGTTGAATATCTATTTGGTGCCTCTACGTGA 4074
Db 155834 TTTGCTGTGCAATTTTGTGAGACATTTGTTGAATATCTATTTGGTGCCTCTACGTGA 155893
QY 4075 TTTTCTTTTAAATCTACTGATATCTGTTCTTAAATTTTCTTCACATATGTTTG 4134
Db 155894 TTTTCTTTTAAATCTACTGATATCTGTTCTTAAATTTTCTTCACATATGTTTG 155953
QY 4135 CCGATATCAACATGATTTTATTAACGTAATTTAAGAAATCTAAGCTAAACCTCAGTTAA 4194
Db 155954 CCGATATCAACATGATTTTATTAACGTAATTTAAGAAATCTAAGCTAAACCTCAGTTAA 156013
QY 4195 GTGCATTTATTTCTTATTAACATAGACCGTGTCTACTCTCAGACACCTCTCTCATATT 4254
|||||

Db 156014 GTGCAATATATTTCTTATACATAGACCGGTCTACTCTCAGACCCCTCCCTCAATT 156073

QY 4255 TTTTTCCTAGCATGTGATGCGCTGATTAACATTTTCATTTGCTTTTATTTCAATA 4314

Db 156074 TTTTTCCTAGCATGTGATGCGCTGATTAACATTTTCATTTGCTTTTATTTCAATA 156133

QY 4315 TGGGAACAATGAGAGTGAACCTTAATATATAGTTGATTAATAAATCATCATTTAGCCTAA 4374

Db 156134 TGGGAACAATGAGAGTGAACCTTAATATATAGTTGATTAATAAATCATCATTTAGCCTAA 156193

QY 4375 TTATTTAGAAAATGCTATTAATAGTACGACATAGAAAATGCTATGCTATTTGT 4434

Db 156194 TTATTTAGAAAATGCTATTAATAGTACGACATAGAAAATGCTATGCTATTTGT 156253

QY 4435 ACCTTTGACGCAATTTTACAGCATTAATGTTGCTATTAATTTAAATAAGTCTTG 4494

Db 156254 ACCTTTGACGCAATTTTACAGCATTAATGTTGCTATTAATTTAAATAAGTCTTG 156313

QY 4495 GGTTCAGAAATACCTTCAAAA 4515

Db 156314 GGTTCAGAAATACCTTCAAAA 156334

RESULT 6
LOCUS BC027108 1496 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, clone IMAGE:5356411, mRNA, partial cds.
ACCESSION BC027108
VERSION BC027108.1 GI:20071983

KEYWORDS

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 1496)
Strausberg, R.
Direct Submission
Submitted (04-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amgdbcm.tmc.edu
Gunnarone, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINTL at: <http://image.lnl.gov>
Series: IRAC Plate: 54 Row: d Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES

location/Qualifiers
1..1496
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:5356411"
/tissue_type="Mammary tumor, C3(1)-Tag model, Infiltrating
ductal carcinoma, 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

CDS
1..680
/codon_start=3
/product="Unknown (protein for IMAGE:5356411)"
/protein_id="AAH27108.1"
/db_xref="GI:20071984"
/translation="PTTLVODMAQILANALLAEFPEITVENLPTLYRFORPLLLFSG
GSINQYRNITLALVRKQDLSFPEKMWLNKTPYGRSILKAYFERPLLPOLLVNL
HSGGVYAFPSOSVTEOSLYLMLHLQAGLENPITYLSAGEMKPLPAFDLNMMDA
PTSOAPTRKVLVECKEAEVQSEAEIQPDDBRSTAREPEVBLIRKRWNNANPKDQEP
FHHDKEI"
BASE COUNT 435 a 329 c 322 g 410 t
ORIGIN

Query Match 9.4%; Score 426.4; DB 10; Length 1496;
Best Local Similarity 65.4%; Pred. No. 1.2e-75;
Matches 756; Conservative 0; Mismatches 326; Indels 74; Gaps 6;

QY 2101 TAGCTAGCACAGATGACAGACATAGTTCATTAATAACAGATGACACTAGGAATGT 2160
1 TAGACACCAACCTGTGTAAGACATGGCTCAGATCTAGCAATGCAATGCACTAGGAAGCAT 60
Db 2161 TTCCGGAATTCACCTGTGGAATCTTCCAGTTATTTCAGACTTCAGAAACATTTTGA 2220
61 TTCCAGAAATCACTGTGGAATCTTCCAGATATTTAAGATTTCAGAGACCGCTACTTC 120
QY 2221 TTTTGTTCAGTATGTCGACGTAAATCCCAATATTAATAAAGCAATATGACACGTGTA 2280
121 TCTTGTTTAGTGTGACACATTAACCTCAGTAGAGAACAGATCTGGACATGTGTA 180
QY 2281 AGCAGAAATACCTTGATTCATTTACTCCATGCTGCTTAATCTAAAGAAATCTCCAGTGG 2340
181 GACAGAGACAGCGATGCTGTTTACCCCTTGCGTGAATATCGAAGAACCCCGCTGG 240
QY 2341 GGAAGAGAACTTTCGCGGATATTTGATCTCTGCTCCCTCTCTCTCTCTCTCTCTGG 2400
241 GGAGGGGCACTTGAAGGATATTTTGGCGCTGCGACACACTTCTCAGCTGCTGTTGG 300
QY 2401 TGAATCTGATTCAGGTGCGCAAGTATTTGCAATTCCTTCAGACCGGCTATTAATTGAAG 2460
301 TGAATCTGATTCAGGTGCGCAAGTATTTGCAATTCCTTCAGACCGGCTATTAATTGAAG 360
Db 2461 AAAACCTTATTTGTCGCTGAAGAAATTTAGAGACAGACTAGAAATCATATCATCAATTT 2520
361 AAAAGCTTGTGTTGGCTGGAACAATTTACAGCGGACTAGAAATCCCATCAGCTGT 420
QY 2521 TACCTGCTTAAGAATGGAACCTCTCTTCAGCTTATGATTTTCTAAGTATGATGATG 2580
421 TGTCTGCCAGGAATGGAACCT 480
QY 2581 CCGCAACATCTCAAGCTGACCTAGGAAGTTCCCAAGTGTGAAGAAAGACAGATGTGC 2640
481 CTCTCATCTCCCAAGCTCCACAGAGAAAGTTCTAGAGTGTAGAGAAAGACAGAGTGC 540
QY 2641 AGAAGATGATTAAGAACCAACATGAATTAATCGGCAGCTCAGAAAAGACAGATTGAAA 2700
541 AGAAGATGATTAAGAACCAACATGAATTAATCGGCAGCTCAGAAAAGACAGATTGAAA 600
Db 2701 CTCTGAGATTAAGATTTGGAATAGAGTAATTTGTTTAAAGACAGAAAATCATTTTA 2760
601 TCTGAGATTAAGATTTGGAATAGAGTAATTTGTTTAAAGACAGAAAATCATTTTA 660
QY 2761 GACGTGATTAAGATTTGGAATAGAGTAATTTGTTTAAAGACAGAAAATCATTTTA 2820
661 ACCAGACAGAGATTAAGAGCGG-----GCCCTGTCGAGATGTC 699
Db 2821 AAATTTTTCGATGATAGACTTAATTTTCTTAAAGATATATTAATCATTTTC 2880
700 CTCTCTGTCGACAGAAACCGAGATCAATCTGTAAGAAAGACAGACTTAATATATGTC 759
QY 2881 AAGTTTCGAGACTAGTCCATCAATAGAAATTAATTAATGTCACATTTTATTTTAA 2940
760 ATCTCT--AGACAGAGTCAATGAAGGCTGTGTATCAAAATGAC-----ATGTAA 810

OY	2941	ATTTCCTGTACTCATTTAAACCAAGTAAGTACGAGCGCCAAATATTTTATATTT	3000
Db	811	ATTTCCTGTACTCATTTAAACCAAGTAAACCGTAACCGCAGAGA-----	851
OY	3001	ACTTTTCAOCCGAGTAGTATATACCATAATGCAATATATGAATATTTAATGAGATATTT	3060
Db	852	-----TTAACCCGACAGCGCTTAATTTACTACGAAATAGTTGCAT----	891
OY	3061	TACATCCTTTTTTGTAACAAGTCTCTTAATGCAATATTTTATTTACTATCTGATTT	3120
OY	892	---TTACCTTTTAAAGTGAAGCTTCAAATAACAGTCC-ATTTCACACTATATATGCCCTTT	946
Db	3121	CCTACTTCCGAGTAGCACCATATTTCAAGTGTTCATTGGCCACATGTGCGCTGATCACTGT	3180
Db	947	CCTGACTAAGAGAGATCCGTTTTTGAGGGCTTATTGGCATAGTGGTGATCAGACGTTGT	1006
OY	3181	ATTGACACAGTTAGTACATACAGAAAACTAGCATATTTACTTACTTGTAGCCATGATTT	3240
Db	1007	AATGGGCGAGTTACATACAGACAAACACCAAGTGTCTTGACCTGGGTCTAGCACTCTGT	1066
OY	3241	CTATTGGATTAAAT 3256	
Db	1067	GCAATTGAATCAAT 1082	
RESULT 7			
AC021131			
LOCUS			
DEFINITION	204606 bp DNA linear HTG 07-JUL-2000		
SEQUENCE	Homo sapiens chromosome 14 clone RP11-382I20, WORKING DRAFT		
ACCESSION	AC021131.5 GI:8569986		
VERSION	HTG: HTGS_PHASE1; HTGS_DRAFT.		
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Waterston,R.H.		
JOURNAL	The sequence of Homo sapiens clone		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 204606)		
TITLE	Waterston,R.H.		
JOURNAL	Direct Submission		
AUTHORS	Submitted (14-JAN-2000) Genome Sequencing Center, Washington		
TITLE	MO 63108, USA		
JOURNAL	On Jun 16, 2000 this sequence version replaced gi:754281.		
COMMENT			
	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site:http://genome.wustl.edu/gsc/index.shtml		
	Project Information		
	Center project name: H.NH0382120		
	Summary Statistics		
	Sequencing vector: M13; 100%		
	Sequencing vector: plasmid; 0%		
	Chemistry: Dye-primer ET; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 0% of reads		
	Assembly program: Phrap; version 0.990319		
	Consensus quality: 192295 bases at least Q40		
	Consensus quality: 195796 bases at least Q30		
	Insert size: 209000; agarose-fp		
	Quality coverage: 4.50 in Q20 bases; agarose-fp		
	Quality coverage: 4.58 in Q20 bases; sum-of-contigs		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 27 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		

	* runs of N, but the exact sizes of the gaps are unknown
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	1 1259: contig of 1259 bp in length
*	1260 1359: gap of unknown length
*	1360 2526: contig of 1167 bp in length
*	2527 2626: gap of unknown length
*	2627 4233: contig of 1607 bp in length
*	4234 4333: gap of unknown length
*	4334 6059: contig of 1726 bp in length
*	6060 6159: gap of unknown length
*	6160 8588: contig of 2429 bp in length
*	8589 8688: gap of unknown length
*	8689 11632: contig of 2944 bp in length
*	11633 11732: gap of unknown length
*	11733 14000: contig of 2268 bp in length
*	14001 14100: gap of unknown length
*	14101 16972: contig of 2872 bp in length
*	16973 17072: gap of unknown length
*	17073 21089: contig of 4017 bp in length
*	21090 21189: gap of unknown length
*	21190 25902: contig of 4713 bp in length
*	25903 26002: gap of unknown length
*	26003 29817: contig of 3815 bp in length
*	29818 29918: gap of unknown length
*	29918 33616: contig of 3659 bp in length
*	33617 33716: gap of unknown length
*	33717 37223: contig of 3507 bp in length
*	37224 37323: gap of unknown length
*	37324 42725: contig of 5402 bp in length
*	42726 42825: gap of unknown length
*	42826 48145: contig of 5320 bp in length
*	48146 48245: gap of unknown length
*	48246 52585: contig of 4340 bp in length
*	52586 52685: gap of unknown length
*	52686 58396: contig of 5711 bp in length
*	58397 58496: gap of unknown length
*	58497 64865: contig of 6369 bp in length
*	64866 64965: gap of unknown length
*	64966 70722: contig of 5757 bp in length
*	70723 70822: gap of unknown length
*	70823 77743: contig of 6921 bp in length
*	77744 77843: gap of unknown length
*	77844 88107: contig of 10264 bp in length
*	88108 88207: gap of unknown length
*	88208 99343: contig of 11136 bp in length
*	99344 99443: gap of unknown length
*	99444 112627: contig of 13184 bp in length
*	112628 112727: gap of unknown length
*	112728 125106: contig of 12279 bp in length
*	125107 125106: gap of unknown length
*	125107 136350: contig of 11244 bp in length
*	136351 136450: gap of unknown length
*	136451 163903: contig of 27453 bp in length
*	163904 164003: gap of unknown length
*	164004 204606: contig of 40603 bp in length.
FEATURES	
source	1. .204606
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="14"
	/clone="RP11-382120"
misc_feature	1. 1259
	/note="assembly_name:Contig13"
misc_feature	1360. 2526
	/note="assembly_name:Contig14"
misc_feature	2627. 4233
	/note="assembly_name:Contig15"
misc_feature	4334. 6059
	/note="assembly_name:Contig16"
misc_feature	6160. 8588
	/note="assembly_name:Contig17"

```

misc_feature      8689..11632
                    /note="assembly_name:Contig18"
misc_feature      11733..14000
                    /note="assembly_name:Contig19"
misc_feature      14101..16972
                    /note="assembly_name:Contig20"
misc_feature      17073..21089
                    /note="assembly_name:Contig21"
misc_feature      21190..25902
                    /note="assembly_name:Contig22"
misc_feature      26003..29817
                    /note="assembly_name:Contig23"
misc_feature      29918..33616
                    /note="assembly_name:Contig24"
misc_feature      33717..37223
                    /note="assembly_name:Contig25"
misc_feature      37324..42725
                    /note="assembly_name:Contig26"
misc_feature      42826..48145
                    /note="assembly_name:Contig27"
misc_feature      48246..52585
                    /note="assembly_name:Contig28"
misc_feature      52686..58396
                    /note="assembly_name:Contig29"
misc_feature      58497..64865
                    /note="assembly_name:Contig30"
misc_feature      64966..70722
                    /note="assembly_name:Contig31"
misc_feature      70823..77743
                    /note="assembly_name:Contig32"
misc_feature      77844..88107
                    /note="assembly_name:Contig33"
misc_feature      88208..99343
                    /note="assembly_name:Contig34"
misc_feature      99444..112627
                    /note="assembly_name:Contig35"
misc_feature      112728..125006
                    /note="assembly_name:Contig36
                    clone_end:SP6
                    vector_side:right"
misc_feature      125107..136350
                    /note="assembly_name:Contig37"
misc_feature      136451..163903
                    /note="assembly_name:Contig38"
misc_feature      164004..204606
                    /note="assembly_name:Contig39"
BASE COUNT      64151 a 38935 c 39288 g 59611 t 2621 others
ORIGIN
Query Match      6.8%; Score 307.2; DB 2; Length 204606;
Best Local Similarity 68.3%; Pred. No. 1.5e-51;
Matches 358; Conservative 0; Mismatches 165; Indels 3; Gaps 1;

```

```

QY 4292 TTCATTGCTTTTATTTCTAATATGGAACAATGAGTAGAAGTCTAATATAGTTGTA 4351
    |||||||
DB 29999 TTCATTGCTTTTATTTCTAATATGGAACAATGAGTAGAAGTCTAATATAGTTGTA 30058
QY 4352 GTAATAAACATCATTTAGGCTTAATTTAGAAAATGCTAATTAAGTACGACACATAGAA 4411
    |||||||
DB 30059 GTAATAAACATCATTTAGGCTTAATTTAGAAAATGCTAATTAAGTACGACACATAGAA 30118
QY 4412 ACATGAATGCTTAGTACATTTGACCTTTGTCAGCAATTTTACAGCTCAATGTTTGT 4471
    |||||||
DB 30119 ACATGAATGCTTAGTACATTTGACCTTTGTCAGCAATTTTACAGCTCAATGTTTGT 30178
QY 4472 CATTAATTTAATTAAGTGTGCTGGTTTCAGAAATACCTTCAAA 4515
    |||||||
DB 30179 CATTAATTTAATTAAGTGTGCTGGTTTCAGAAATACCTTCAAA 30222

```

```

RESULT 8
LOCUS      AX079596 547 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 340 from Patent WO0107611.
ACCESSION  AX079596
VERSION     AX079596.1 GI:13159156
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 547)
AUTHORS     Baker, K.P., Goddard, A. and Wood, W.I.
TITLE       Human polypeptides and methods for the use thereof
JOURNALS    Patent: WO 0107611-A 340 01-FEB-2001;
            Genentech, Inc. (US)
FEATURES
Source      Location/Qualifiers
            1..547
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  155 a 117 c 131 g 139 t 5 others
ORIGIN
Query Match      6.4%; Score 290; DB 6; Length 547;
Best Local Similarity 88.8%; Pred. No. 4e-48;
Matches 348; Conservative 0; Mismatches 37; Indels 7; Gaps 3;

```

```

QY 111 CCTCTCCGTCGCGCGCGGCAACTGTTGCGCGCGCTCGGGAACGCCAGGT 170
    |||||||
DB 114 CCTCTCTCCGTCGCGCGGCAACTGTTGCGCGCGCTCGGGAACGCCAGGT 172
QY 171 CCGCGCGCGAGTCCGCGGAGATACATAGATCATCGTAGAAGTCTTGAAGTTG 230
    |||||||
DB 173 CCGCGCGCGAGTCCGCGGAGATACATAGATCATCGTAGAAGTCTTGAAGTTG 228
QY 231 TTCAGAAAAATTTGAAGTAGCAAAATAGAAAATTAAGAAATTAACAGCATACAGAG 290
    |||||||
DB 229 TTCAGAAAAATTTGAAGTAGCAAAATAGAAAATTAAGAAATTAACAGCATACAGAG 288
QY 291 ACAGCATGAAGTGTGCTTAGGAACAGACAGACAGTGAAGAAACAGCAAAATTC 350
    |||||||
DB 289 --CAGCATGAAGTGTGCTTAGGAACAGACAGACAGTGAAGAAACAGCAAAATTC 346
QY 351 GCTCAGATACACAGTCAAGTGAATGTTTCCGCTCAATGCTTGAAGTGGAGT 410
    |||||||
DB 347 GCTCAGATACACAGTCAAGTGAATGTTTCCGCTCAATGCTTGAAGTGGAGT 406
QY 411 TCTTTGTCATATATGTCATTTTTCATCATGCAACAGTAACCTTTTACCAAGTGA 470
    |||||||
DB 407 TCTTTGTCATATATGTCATTTTTCATCATGCAACAGTAACCTTTTACCAAGTGA 466
QY 471 COTCAGAAATATTTAGTACATTTGCAACAGG 502
    |||||||
DB 467 COTCAGAAATATTTAGTACATTTGCAACAGG 498

```


KEYWORDS HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
TITLE 1 (bases 1 to 203610)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
JOURNAL Mus musculus, clone RP23-248E1
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 203610)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 203610)
AUTHORS Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choquel, Y., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 21, 2002 this sequence version replaced gi:17061655. All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
----- Project Information
Contact: sequence_submissions@genome.wi.mit.edu
Center project name: L19034
Center clone name: 248_E1
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 194595 bases at least Q40
Consensus quality: 198128 bases at least Q30
Consensus quality: 199686 bases at least Q20
Insert size: 204000; agarose-fp
Insert size: 200810; sum-of-contigs
Quality coverage: 6.3 in Q20 bases; agarose-fp
Quality coverage: 6.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1366: contig of 1366 bp in length
* 1367 1466: gap of 100 bp
* 1467 2392: contig of 926 bp in length
* 2393 2492: gap of 100 bp
* 2493 3944: contig of 1452 bp in length
* 3945 4044: gap of 100 bp
* 4045 5485: contig of 1441 bp in length
* 5486 5585: gap of 100 bp
* 5586 6815: contig of 1230 bp in length
* 6816 6915: gap of 100 bp
* 6916 8703: contig of 1788 bp in length
* 8704 8803: gap of 100 bp
* 8804 10621: contig of 1818 bp in length
* 10622 10721: gap of 100 bp
* 10722 13000: contig of 2279 bp in length
* 13001 13100: gap of 100 bp
* 13101 15018: contig of 1918 bp in length
* 15019 15118: gap of 100 bp
* 15119 17273: contig of 2155 bp in length
* 17274 17373: gap of 100 bp
* 17374 21030: contig of 3657 bp in length
* 21031 21130: gap of 100 bp
* 21131 24393: contig of 3263 bp in length
* 24394 24493: gap of 100 bp
* 24494 27284: contig of 2791 bp in length
* 27285 27384: gap of 100 bp
* 27385 53071: contig of 25687 bp in length
* 53072 53171: gap of 100 bp
* 53172 56584: contig of 3413 bp in length
* 56585 56684: gap of 100 bp
* 56685 63948: contig of 7264 bp in length
* 63949 64048: gap of 100 bp
* 64049 71335: contig of 7267 bp in length
* 71336 71435: gap of 100 bp
* 71436 78031: contig of 6596 bp in length
* 78032 78131: gap of 100 bp
* 78132 86041: contig of 7910 bp in length
* 86042 86141: gap of 100 bp
* 86142 97779: contig of 11638 bp in length
* 97780 97879: gap of 100 bp
* 97880 106798: contig of 8919 bp in length
* 106799 106898: gap of 100 bp
* 106899 120969: contig of 14071 bp in length
* 120970 121069: gap of 100 bp
* 121070 129692: contig of 8623 bp in length
* 129693 129792: gap of 100 bp
* 129793 141332: contig of 11540 bp in length
* 141333 141432: gap of 100 bp
* 141433 151350: contig of 9918 bp in length
* 151351 151450: gap of 100 bp
* 151451 163581: contig of 12131 bp in length
* 163582 163681: gap of 100 bp
* 163682 178518: contig of 14837 bp in length
* 178519 178618: gap of 100 bp
* 178619 201081: contig of 22463 bp in length
* 201082 201181: gap of 100 bp
* 201182 203610: contig of 2429 bp in length.


```
repeat_region 14206..14251
/note="23 copies 2 mer tc 74% conserved"
repeat_region 15594..15633
/note="20 copies 2 mer gt 90% conserved"
repeat_region 15814..16129
/note="L1MB3 repeat: matches 5851..6181 of consensus"
repeat_region 16968..17005
/note="19 copies 2 mer ca 100% conserved"
misc_feature complement(1752..18167)
/note="match: GSS A0065232"
repeat_region 18176..18830
/note="L1MD3 repeat: matches 7082..7740 of consensus"
repeat_region 19980..20318
/note="L1LAI repeat: matches 1..337 of consensus"
repeat_region 20320..20418
/note="L1R12 repeat: matches 572..670 of consensus"
repeat_region 20421..20512
/note="L1R30 repeat: matches 551..643 of consensus"
repeat_region 20513..20736
/note="L1R12 repeat: matches 14..1440 of consensus"
misc_feature complement(20664..20987)
/note="match: GSS B33497"
repeat_region 20807..20985
/note="MIR repeat: matches 65..245 of consensus"
repeat_region 20988..21011
/note="12 copies 2 mer tt 96% conserved"
repeat_region 21201..21393
/note="MIR repeat: matches 57..262 of consensus"
repeat_region 24587..24675
/note="L1R23 repeat: matches 304..395 of consensus"
repeat_region 24902..25190
/note="Alu repeat: matches 1..292 of consensus"
repeat_region 25655..25754
/note="50 copies 2 mer ta 85% conserved"
repeat_region 25991..26294
/note="Alu repeat: matches 1..312 of consensus"
repeat_region 26461..26661
/note="MIR repeat: matches 19..228 of consensus"
misc_feature 28740..29322
/note="match: GSS A0348225"
repeat_region 29629..29699
/note="MIR repeat: matches 59..131 of consensus"
misc_feature complement(31515..31877)
/note="match: GSS A0100296"
repeat_region 32772..32846
/note="L2 repeat: matches 2675..2750 of consensus"
repeat_region 32825..32951
/note="L2 repeat: matches 1729..1860 of consensus"
repeat_region 33291..33572
/note="Charles repeat: matches 7..291 of consensus"
repeat_region 33577..33745
/note="Charles repeat: matches 581..743 of consensus"
repeat_region 33779..35181
/note="Charles repeat: matches 901..2297 of consensus"
repeat_region 35187..35513
/note="L1MA10 repeat: matches 5994..6316 of consensus"
misc_feature 35272..35512
/note="match: STR 224589; genomic marker D6S470"
repeat_region 35348..35379
/note="16 copies of ca 100% conserved; differs from
224589"
repeat_region 35522..35782
/note="Charles repeat: matches 2289..2574 of consensus"
repeat_region 35921..35944
/note="12 copies 2 mer aa 96% conserved"
repeat_region 37469..38001
/note="L2 repeat: matches 2089..2627 of consensus"
repeat_region 38068..38375
/note="MER63B repeat: matches 8..335 of consensus"
repeat_region 38376..38542
/note="Alu repeat: matches 133..299 of consensus"
repeat_region 38742..39153
/note="L2 repeat: matches 2310..2705 of consensus"

repeat_region 39184..39447
/note="Alu repeat: matches 38..302 of consensus"
repeat_region 39451..39832
/note="L1P repeat: matches 558..961 of consensus"
repeat_region 39935..40111
/note="Alu repeat: matches 132..308 of consensus"
repeat_region 40114..40252
/note="L2 repeat: matches 1555..1693 of consensus"
repeat_region 40486..40575
/note="L2 repeat: matches 2598..2727 of consensus"
repeat_region 41332..41498
/note="MIR repeat: matches 55..228 of consensus"
repeat_region 41558..41862
/note="Alu repeat: matches 1..305 of consensus"
repeat_region 43659..43959
/note="Alu repeat: matches 1..305 of consensus"
repeat_region 44272..44367
/note="Alu repeat: matches 1..305 of consensus"
repeat_region 45065..45299
/note="MIR repeat: matches 47..140 of consensus"
repeat_region 45488..45925
/note="MIR repeat: matches 1..258 of consensus"
repeat_region 45926..45977
/note="L1P repeat: matches 5706..6143 of consensus"
misc_feature 46301..46562
/note="match: GSS A0167854"
repeat_region 46563..46893
/note="MER2 repeat: matches 1..345 of consensus"
repeat_region 46903..46936
/note="17 copies 2 mer tg 91% conserved"
repeat_region 47565..47793
/note="L2 repeat: matches 1901..2131 of consensus"
repeat_region 48465..48620
/note="L1 repeat: matches 245..410 of consensus"
repeat_region 49388..49548
/note="MIR repeat: matches 33..198 of consensus"
repeat_region 49549..49829
/note="TIGER1 repeat: matches 2..273 of consensus"
repeat_region 49830..50123
/note="Alu repeat: matches 16..309 of consensus"
repeat_region 50124..52184
/note="TIGER1 repeat: matches 273..2417 of consensus"
repeat_region 52185..52247
/note="MIR repeat: matches 198..259 of consensus"
repeat_region 52873..53187
/note="MER2 repeat: matches 15..342 of consensus"
misc_feature complement(52938..53361)
/note="match: GSS A0185926"
repeat_region 54150..54465
/note="L1MA7 repeat: matches 5963..6288 of consensus"

Query Match 2.7%; Score 124; DB 9; Length 94749;
Best Local Similarity 55.9%; Pred. No. 1.4e-14;
Matches 378; Conservative 0; Mismatches 260; Indels 38; Gaps 6;
```



```
RESULT 15
AC012619/c AC012619 89196 bp DNA linear PRI 31-AUG-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2022K1, complete sequence.
DEFINITION AC012619
ACCESSION AC012619
VERSION AC012619.6 GI:15383776
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 89196)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 89196)
DOE Joint Genome Institute.
Direct Submission
3 (bases 1 to 89196)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 89196)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (31-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Aug 31, 2001 this sequence version replaced gi:7711556.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence:
Estimated Total Number of Errors is 0.2.
Location/Qualifiers
1. 89196
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2022K1"
BASE COUNT 30551 a 17310 c 16123 g 25212 t
ORIGIN
Query Match 2.5%; Score 111; DB 9; Length 89196;
Best Local Similarity 64.2%; Pred.No.5.7e-12;
Matches 219; Conservative 0; Mismatches 110; Indels 12; Gaps 3;
QY 2915 AATATAGTCACATATTTTAAATTTCTAGTACTACTTAACAAGTAAAGT 2974
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20986 AATGTGAGCCACATACATATTTAAATTTCTAGTAGCACATTAACCTAAAGG 20927
QY 2975 GAGCAGGCGAANAATTTGATATTTCTTCCAGTAGTATACCCAAATATGCGAA 3034
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20926 AAGCAAGTGAATTAATTTTAAATTAATTAACCAATATATCGAATATTTGTTCA 20867
QY 3035 ATATAGA-----AATTAATGAGATTTTACATCTCTTTTGTACCAAGTC 3084
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20866 ACGTGTAAATCAACATAAATTAATAGAGATATTTTACAT-TTTTTCATGCTAAATCT 20808
QY 3085 TCTAATGCGAGTACATATTTTACTTACGATTTCTTACTTCGAGTAGCCATATTTTC 3144
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20807 TTGAATCCGGTGTGTATCTTAATTTACGGCTTATCTCATTTGGACTTGACACATTCG 20748
QY 3145 AAGTGTC-ATTGCCACATGTGGCTGTGACTACTGATTTGACAGTTCACTAGACACA 3203
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20747 AAGTGCTCAATAGCTATGTGTAGCTATTGACTACTTATTTGGACAGTGCATTTCTAGACC 20688
QY 3204 AAACCTAGCATATTAATTAATTTAGTTCTAGCATATTTCTAT 3244
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20687 TCAAGAACAGAGATTATTTTATCATGTATGATGAATAT 20647
```



```

RESULT 2
US-07-906-871-15/c
: Sequence 15 Application US/07906871
: Patent No. 5340739
: GENERAL INFORMATION:
: APPLICANT: Stevens, Richard L.
: APPLICANT: Avraham, Shalom
: TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
: TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLIXIN AND USES
: TITLE OF INVENTION: THEREOF
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1225 Connecticut Avenue, N.W., Suite 300
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/906,871
: FILING DATE: 19920103
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/816,289
: FILING DATE: 03 JAN 1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/635,544
: FILING DATE: 18-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US89/03051
: FILING DATE: 13-JUL-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/224,035
: FILING DATE: 13-JUL-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Cimbalia, Michele A
: REGISTRATION NUMBER: 33,851
: REFERENCE/DOCKET NUMBER: 0627,2830004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)833-7533
: TELEFAX: (202)833-8716
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 17327 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: exon
: LOCATION: 621..753
: FEATURE:
: NAME/KEY: intron
: LOCATION: 754..9596
: FEATURE:
: NAME/KEY: exon
: LOCATION: 9597..9744
: FEATURE:
: NAME/KEY: intron
: LOCATION: 9745..16396
: FEATURE:
: NAME/KEY: exon
: LOCATION: 16397..17327
: US-07-906-871-15
Query Match 2.0%; Score 91.4; DB 1; Length 17327;
Best Local Similarity 58.9%; Pired. No. 5.3e-13;

```

Matches	195	Conservative	0	Mismatches	131	Indels	5	Gaps	2
QY	2878	TTCAAGTTTGCAGACTAGTGCATTCGCAATAGAAATTAATTAATTAAGTCACATATTTTATTT							2937
Db	5964	TTGTACTGCTCTGTAGACCAACACTTTTGCCAAATGAAACACAAAGCAACACATGTGGAGATT							5905
QY	2938	AAAAATTTTCTACTACTACATATTAACAAAGTAAAGTGAGCGGGCAAAATTAATTTGAT							2997
Db	5904	TAAATGTTCCAGCGGCACATATTTTAAAGTAAAGAAACAGGCGGAAATTAATTTTAT							5845
QY	2998	A----TTACTTTTCACCCAGTAGTATATCCCAAAATAGCGAAATATAGAAATTAATTA							3053
Db	5844	TAACTTAACATATTCOAATATATTTGCTCTTAAACATATCATCATATGTAAGAAAGTATTA							5785
QY	3054	GATATTTTATACATCTTTTGTGACCAAGCTCTTCTAAATGACAGTACATATTTTATCTAC							3113
Db	5784	GATATTTTTCATCTTTCTTCTCATCTAAGGCTCCAAATCTGTATATTTTGTGATTCAC							5725
QY	3114	TGCATTTCTTACTTCCAGTAGTACCAATTTTCAAGTGTTCA--TTGGCACATGTGGCGTGG							3172
Db	5724	AACCTGTGTCCTCAATTTTGCCCTTGCCACATTTTCAAGTGTTCAAGTACGACAGTGA							5665
QY	3173	ACTACTGTATTTGGACAGTTTCAGTACTAGTCA							3203
Db	5664	GCTACCGTACTGAGACAGACAGCGCCTAGCA							5634
RESULT 3									
		US-09-784-316-3							
		; Sequence 3, Application US/09784316							
		; Patent No. 6461843							
		; GENERAL INFORMATION:							
		; APPLICANT: WEI, Ming-Hui et al.							
		; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC							
		; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES							
		; TITLE OF INVENTION: THEREOF							
		; FILE REFERENCE: CLO01139							
		; CURRENT APPLICATION NUMBER: US/09/784,316							
		; CURRENT FILING DATE: 2001-02-16							
		; NUMBER OF SEQ ID NOS: 5							
		; SOFTWARE: FastSeq for Windows Version 4.0							
		; SEQ ID NO 3							
		; LENGTH: 65042							
		; TYPE: DNA							
		; ORGANISM: Human							
		; FEATURE:							
		; NAME/KEY: misc_feature							
		; LOCATION: (1)...(65042)							
		; OTHER INFORMATION: n = A,T,C or G							
		US-09-784-316-3							
Query Match									
		1.7%;	Score 74.8;	DB 4;	Length 65042;				
		Best Local Similarity	58.8%;	Pred. No. 1.9e-08;					
		Matches 188;	Conservative 0;	Mismatches 122;	Indels				


```

; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

```

```
; NAME/KEY: unsure
; LOCATION: 7427
; OTHER INFORMATION: unknown
```


QY	2824	CAATTAATTTTAAATTTTCTAGTAACATAAACAAGTAAAGTGAAGGGC	2893
Db	154963	CACATAGTAATTTTAAATTTTAAATGGCACATTTTA	1549110
		-----	AAAGTGAATAAAGA
QY	2884	AAAATTAATTTGATATACCTTTCCACCCAGTAGTATACCAAAATATGCGAAATATATGAAA	3043
Db	154909	AATTTATTTTAAATGATTTGAAATCAGGTACACCAAAATATGTTTCAACAAAGATTTAAT	1548550
QY	3044	TTATTAATGAGATATTTTACATCCCTTTTGTACCAAGCTTCTAATAATGACAGTACATATT	3103
Db	154849	ATTAAATTAATGAGATTTTCTTTGTATTTTAACTATGTTTGAATATGCGGTGTATT	1547901

[illegible]

```

OY      3044  TAAATTAAGATGATTTTATCTTATCTTTTGGACCAAGCTCTCTAAATGACATGACATATT 3103
          |||      |||      |||      |||      |||      |||      |||      |||
DB 154849  ATTAATAATATTAGATTTCATCTTTGTTATTTACTAGTCTTTGAAATCTGGTGTTATT 154790
OY      3104  TTACTACTACTGATTTCTTACTTCGAGTAGACATATTTCAAGTG-TTCATTGCCACAT 3162
          |||      |||      |||      |||      |||      |||      |||      |||
DB 154789  TTACACTTAAGACACATCATACAGTTTGGATAGACACATTTCCAAATGCTTAATTAATCAGAT 154730
OY      3163  GTGGCTGTGACTGATCTGATTTGGACAGTTTCAG 3194
          |||      |||      |||      |||      |||      |||      |||      |||
Db 154729  ATGTTAGTGCGCACTATCTTGACAGGACAG 154698

RESULT 9
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Rudd, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724.394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match 1.4%; Score 65.2; DB 2; Length 246240;
Best Local Similarity 57.7%; Pred. No. 1.1e-05;
Matches 157; Conservative 0; Mismatches 108; Indels 7; Gaps 2;

OY      2924  CACATATTTTATTTAAATTTCTGTGTAAGTAAACAAGTAAAGTGACAGGCG 2983
          |||||      |||      |||||      |||      |||||      |||      |||||      |||
Db 154963  CACATATGCTAAATTTTAAATTTTATATGAGGCACATTTTA-----AAAAGTAAAAAATA 154910
OY      2984  AAAATAAATTTTGATTTTACTTTTCACCCAGTGTAGTACCCAAATAATAGCAAAATATAGAAA 3043

```

[illegible]

US-09-641-638-23/c
: Sequence 23, Application US/09641638
: Patent No. 6432648
: GENERAL INFORMATION:
: APPLICANT: Blumenfeld, Marta
: APPLICANT: Bougueleret, Lydie
: APPLICANT: Chumakov, Ilya
: APPLICANT: Cohen, Annick
: TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
: TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
: FILE REFERENCE: GENSET.051CPI
: CURRENT APPLICATION NUMBER: US/09/641,638
: PRIOR FILING DATE: 2000-08-16
: PRIOR APPLICATION NUMBER: US 09/502,330
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: US 60/133,200
: PRIOR FILING DATE: 1999-05-07
: PRIOR APPLICATION NUMBER: US 09/275,267
: PRIOR FILING DATE: 1999-03-23
: PRIOR APPLICATION NUMBER: US 60/119,917
: PRIOR FILING DATE: 1999-02-12
: NUMBER OF SEQ ID NOS: 1304
: SOFTWARE: Patent.pm
: SEQ ID NO 23
: LENGTH: 955
: TYPE: DNA
: ORGANISM: Homo Sapiens
: FEATURE:
: NAME/KEY: allele
: LOCATION: 478
: OTHER INFORMATION: 12-628-306 : polymorphic base G or A
: NAME/KEY: misc_binding
: LOCATION: 479..497
: OTHER INFORMATION: 12-628-306.misl, complement
: NAME/KEY: misc_binding
: LOCATION: 458..477
: OTHER INFORMATION: 12-628-306.mis2, potential
: NAME/KEY: primer_bind
: LOCATION: 764..782
: OTHER INFORMATION: upstream amplification primer, complement
: NAME/KEY: primer_bind
: LOCATION: 266..286
: OTHER INFORMATION: downstream amplification primer
: NAME/KEY: misc_binding
: LOCATION: 466..490
: OTHER INFORMATION: 12-628-306 potential probe
: US-09-641-638-23

Query Match 1.4%; Score 63.4; DB 4; Length 955;
Best Local Similarity 64.9%; Pred. No. 1.5e-06;
Matches 111; Conservative 0; Mismatches 56; Indels 4; Gaps 1;

QY 3043 ATTATTAAGAGATATTACATCC---TTTTTGTACCAAGTCTTAATGCAGTAC 3098
|||||
DB 899 ATTATTAAGAGATATTACATCC---TTTTTGTATTAACCTTCAAACTCCGGTGT 840
|||||

QY 3099 AATTATTAAGAGATATTACATCC---TTTTTGTACCAAGTCTTAATGCAGTAC 3158
|||||
DB 839 GCAATTTACATTTACAGCAGCTCTCAGTTCAGAGAGAGCTACAGGCTCAGGAGC 780
|||||

QY 3159 ACATGTGGCTGTGACTGATTTGAGACAGTTCAGTACTAGACAAAACT 3209
|||
DB 779 CACACTGACTAGCACTACTGATTTGGGTAGTACAGCTACAGCAAAAT 729
|||||

RESULT 12
US-09-641-638-24/c
: Sequence 24, Application US/09641638
: Patent No. 6432648
: GENERAL INFORMATION:
: APPLICANT: Blumenfeld, Marta
: APPLICANT: Bougueleret, Lydie
: APPLICANT: Chumakov, Ilya

APPLICANT: Cohen, Annick
: TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
: TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
: FILE REFERENCE: GENSET.051CPI
: CURRENT APPLICATION NUMBER: US/09/641,638
: PRIOR FILING DATE: 2000-08-16
: PRIOR APPLICATION NUMBER: US 09/502,330
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: US 60/133,200
: PRIOR FILING DATE: 1999-05-07
: PRIOR APPLICATION NUMBER: US 09/275,267
: PRIOR FILING DATE: 1999-03-23
: PRIOR APPLICATION NUMBER: US 60/119,917
: PRIOR FILING DATE: 1999-02-12
: NUMBER OF SEQ ID NOS: 1304
: SOFTWARE: Patent.pm
: SEQ ID NO 24
: LENGTH: 955
: TYPE: DNA
: ORGANISM: Homo Sapiens
: FEATURE:
: NAME/KEY: allele
: LOCATION: 478
: OTHER INFORMATION: 12-628-311 : polymorphic base T or C
: NAME/KEY: misc_binding
: LOCATION: 458..477
: OTHER INFORMATION: 12-628-311.misl, potential
: NAME/KEY: misc_binding
: LOCATION: 479..498
: OTHER INFORMATION: 12-628-311.mis2, potential complement
: NAME/KEY: primer_bind
: LOCATION: 769..787
: OTHER INFORMATION: upstream amplification primer, complement
: NAME/KEY: primer_bind
: LOCATION: 271..291
: OTHER INFORMATION: downstream amplification primer
: NAME/KEY: misc_binding
: LOCATION: 466..490
: OTHER INFORMATION: 12-628-311 potential probe
: US-09-641-638-24

Query Match 1.4%; Score 63.4; DB 4; Length 955;
Best Local Similarity 64.9%; Pred. No. 1.5e-06;
Matches 111; Conservative 0; Mismatches 56; Indels 4; Gaps 1;

QY 3043 ATTATTAAGAGATATTACATCC---TTTTTGTACCAAGTCTTCAATGCAGTAC 3098
|||||
DB 904 ATTATTAAGAGATATTACATCC---TTTTTGTATTAACCTTCAAACTCCGGTGT 845
|||||

QY 3099 AATTATTAAGAGATATTACATCC---TTTTTGTACCAAGTCTTCAATGCAGTAC 3158
|||||
DB 844 GCAATTTACATTTACAGCAGCTCTCAGTTCAGAGAGAGCTACAGGCTCAGGAGC 785
|||||

QY 3159 ACATGTGGCTGTGACTGATTTGAGACAGTTCAGTACTAGACAAAACT 3209
|||
DB 784 CACACTGACTAGCACTACTGATTTGGGTAGTACAGCTACAGCAAAAT 734
|||||

RESULT 13
US-09-813-133A-3/c
: Sequence 3, Application US/09813133A
: Patent No. 6455294
: GENERAL INFORMATION:
: APPLICANT: GAN, Weiniu et al
: TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: C1001173
: CURRENT APPLICATION NUMBER: US/09/813,133A
: PRIOR FILING DATE: 2001-06-06
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on:      March 22, 2003, 09:31:52 ; Search time 329 Seconds
            (without alignments)
            10668.096 Million cell updates/sec
```

Title:	US-09-847-046-1
Perfect score:	4526
Sequence:	1 ggcgagtcggtgcygctgc.....ccttcaaaaaaaaaa 4526

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      538826 seqs, 387737923 residues
Total number of hits satisfying chosen parameters: 10776522
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

1:  published.Applications_NA.*
2:  /cgn2_6/ptodata/2/pubnpa/US07_PUBCOMB.seq.*
3:  /cgn2_6/ptodata/2/pubnpa/PTCT_NEW_PUB.seq.*
4:  /cgn2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq.*
5:  /cgn2_6/ptodata/2/pubnpa/US06_PUBCOMB.seq.*
6:  /cgn2_6/ptodata/2/pubnpa/US07_NEW_PUB.seq.*
7:  /cgn2_6/ptodata/2/pubnpa/PTCT_PUBCOMB.seq.*
8:  /cgn2_6/ptodata/2/pubnpa/US08_NEW_PUB.seq.*
9:  /cgn2_6/ptodata/2/pubnpa/US08_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubnpa/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubnpa/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubnpa/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubnpa/US60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	4525.6	100.0	4526	10	US-09-847-046-1	Sequence 1, Appl1
2	106	2.3	1423	10	US-09-764-864-358	Sequence 358, App
3	97.8	2.2	6792	9	US-10-121-235-20	Sequence 20, Appl
4	86.8	1.9	31208	10	US-09-852-067-3	Sequence 3, Appl1
5	82.4	1.8	378361	9	US-09-901-136-3	Sequence 3, Appl1
6	81.4	1.8	1083	10	US-09-731-872-239	Sequence 239, App
7	79.8	1.8	546	10	US-09-864-761-11995	Sequence 11995, App
8	79.6	1.8	147309	10	US-09-742-312-3	Sequence 3, Appl1
9	76.2	1.7	34337	10	US-09-741-149-3	Sequence 3, Appl1
10	76	1.7	13605	10	US-09-764-847-3446	Sequence 3446, App
11	74.8	1.7	65042	9	US-10-320-124-3	Sequence 3, Appl1
12	74	1.6	157875	9	US-09-935-464-1	Sequence 1, Appl1
13	72.2	1.6	66686	10	US-09-736-960-86	Sequence 86, Appl1
14	71	1.6	197997	10	US-09-822-246-3	Sequence 3, Appl1
15	70	1.5	12047	10	US-09-764-847-1297	Sequence 1297, App
16	70	1.5	15387	10	US-09-764-847-1301	Sequence 1301, App
17	69.2	1.5	147309	10	US-09-742-312-3	Sequence 3, Appl1
18	68.4	1.5	10935	10	US-09-764-847-1300	Sequence 1300, App
19	68.4	1.5	13715	10	US-09-764-847-1298	Sequence 1298, App

C	20	68.4	1.5	197997	10	US-09-852-246-3	Sequence 3, Appl1
C	21	66.8	1.5	75869	10	US-09-864-883-743	Sequence 243, Appl1
C	22	66	1.5	30352	10	US-09-764-869-1768	Sequence 1768, Appl1
C	23	65.2	1.4	3589	10	US-09-925-297-241	Sequence 241, Appl1
C	24	64.6	1.4	22889	9	US-09-860-670-164	Sequence 164, Appl1
C	25	64	1.4	33206	10	US-09-880-107-2380	Sequence 2380, Appl1
C	26	63.8	1.4	1745	10	US-09-882-103-804-10	Sequence 10, Appl1
C	27	63.8	1.4	3425	10	US-09-764-887-315	Sequence 315, Appl1
C	28	63.8	1.4	6945	10	US-09-764-887-313	Sequence 313, Appl1
C	29	63.8	1.4	6945	10	US-09-764-887-314	Sequence 314, Appl1
C	30	63.2	1.4	2547	10	US-09-925-302-323	Sequence 232, Appl1
C	31	63.2	1.4	15714	10	US-09-764-869-2355	Sequence 2355, Appl1
C	32	62.6	1.4	8807	10	US-09-764-869-1415	Sequence 1415, Appl1
C	33	62.2	1.4	403	9	US-09-946-807-1462	Sequence 1462, Appl1
C	34	62.2	1.4	403	10	US-09-795-668-1462	Sequence 1462, Appl1
C	35	62.2	1.4	403	10	US-09-795-668-1462	Sequence 1462, Appl1
C	36	61.4	1.4	485	9	US-09-796-692-7988	Sequence 7988, Appl1
C	37	60.6	1.3	1125	10	US-09-764-869-2422	Sequence 2422, Appl1
C	38	60.6	1.3	143306	9	US-09-729-920-3	Sequence 3, Appl1
C	39	60.6	1.3	174493	9	US-10-238-709-3	Sequence 3, Appl1
C	40	60.6	1.3	174493	10	US-09-804-471A-3	Sequence 3, Appl1
C	41	58.4	1.3	197496	9	US-09-877-177-10	Sequence 10, Appl1
C	42	58	1.3	7032	10	US-09-764-847-1429	Sequence 1429, Appl1
C	43	58	1.3	31766	10	US-09-765-344-5	Sequence 5, Appl1
C	44	58	1.3	63000	10	US-09-780-172-18	Sequence 18, Appl1
C	45	57.8	1.3	3734	10	US-09-764-653-862	Sequence 862, Appl1

ALIGNMENTS

RESULT 1
US-09-847-046-1

Query Match	100.0%	Score 4525.6	DB 10	Length 4526
Best Local Similarity	100.0%	Pred. No. 0		
Matches 4526	Conservative 0	Mismatches 0	Indels 0	Gaps
QY	1	GCCGAGTCGGTGGCGGCTGCAGGCTGGAGGAGAAAGTGTACGCTTTGACAGGTTGGCG	60	
Db	1	GCGCAGTCGGGTGGCGGCTGCAGGCTGGAGAGGAATAAGTGTACGCCCTTGAGAGTTGGCG	60	
QY	61	AAGTGGTTCAGGCTACCCGGGCTAAGTGTGGCAGGCGCCCGCTTCTTGCTCTCTCCGT	120	
Db	61	AAGGTGTTCCAGGCTACCCGGGCTAAGTGTGGCAGGCGCCCGCTTCTTGCTCTCTCCGT	120	
QY	121	CGGTGGCGGCGGGAACATGTTGGCCGGCGGGCTCTGGGGAACGGGCCAGATCCCGCCGC	180	
Db	121	CGGTGGCGGCGGGAACATGTTGGCCGGCGGGCTCTGGGGAACGGGCCAGATCCCGCCGC	180	

181 AGGTCCCGGCGAGATACATAGATCATAGAGAACTTCTGAACTGTTCAAGAAAA 240
181 AGGTCCCGGCGAGATACATAGATCATAGAGAACTTCTGAACTGTTCAAGAAAA 240
241 ATTGTGAAGTACGCAAAATAGAAAAATTAAGAAATTAACAGCAGATACAGAGACATGGA 300
241 ATTGTGAAGTACGCAAAATAGAAAAATTAAGAAATTAACAGCAGATACAGAGACATGGA 300
301 AGTGTGCTTTAGAAACAGACAGCAGCAGTGAAGAAAAACAGCAAAATCCGCTCAGATAC 360
301 AGTGTGCTTTAGAAACAGACAGCAGCAGTGAAGAAAAACAGCAAAATCCGCTCAGATAC 360
361 AACTGACGCTGATTAATGTTTTCCGGCTTCAATGCTTTAGAGTTGGGATCTCTTTGTCA 420
361 AACTGACGCTGATTAATGTTTTCCGGCTTCAATGCTTTAGAGTTGGGATCTCTTTGTCA 420
421 TAATGTGCTATTTTACATGCCCAACAGTAAACTCTTACAGAACTGAGTCTCAGAAAT 480
421 TAATGTGCTATTTTACATGCCCAACAGTAAACTCTTACAGAACTGAGTCTCAGAAAT 480
481 ATTTAGTACATTTGCAACAGCTGCTTGAAGACATGAATGAGGCTGTTAGACCTCTGAGG 540
481 ATTTAGTACATTTGCAACAGCTGCTTGAAGACATGAATGAGGCTGTTAGACCTCTGAGG 540
541 ACTATGGAATTTGAGTTGCCAAGGTTAATGTGTCAAGAAATATCAAGATCTGTG 600
541 ACTATGGAATTTGAGTTGCCAAGGTTAATGTGTCAAGAAATATCAAGATCTGTG 600
601 GAAAGAAAGAAAGTATGATGAAGCAATATTTATTCAGAGGCAACATATTCCTCAGAAAT 660
601 GAAAGAAAGAAAGTATGATGAAGCAATATTTATTCAGAGGCAACATATTCCTCAGAAAT 660
661 TCCCTACTGACACCTGTTTGTGATGTGAATGCAATGTCGCGCAATGTTCTTGTCTCTTC 720
661 TCCCTACTGACACCTGTTTGTGATGTGAATGCAATGTCGCGCAATGTTCTTGTCTCTTC 720
721 TTTTATGTGAAGTAAATATATTTACCAACCTGGAAGACCTTCAGAACATAGAAAAATGCTC 780
721 TTTTATGTGAAGTAAATATATTTACCAACCTGGAAGACCTTCAGAACATAGAAAAATGCTC 780
781 TGAAGAAAGAAAGCAATATATATTTCTCATATGTAAGAGCCATTTGGAATACAGAGACA 840
781 TGAAGAAAGAAAGCAATATATATTTCTCATATGTAAGAGCCATTTGGAATACAGAGACA 840
841 GAGCAGTACATGGAAGCCGCTTGTGTATGAGTACATACCAATTTGTTTAAACACAG 900
841 GAGCAGTACATGGAAGCCGCTTGTGTATGAGTACATACCAATTTGTTTAAACACAG 900
901 AAATGCCCCCTTTTGAAGATATTTGGCTCTGAGAGATGGAATATGCACATCTCTACTTTT 960
901 AAATGCCCCCTTTTGAAGATATTTGGCTCTGAGAGATGGAATATGCACATCTCTACTTTT 960
961 TTTATTTGTAACTAGTCTTGTGACTGACCCAGCAATGTAGAAAGCACTATATGAAACAGC 1020
961 TTTATTTGTAACTAGTCTTGTGACTGACCCAGCAATGTAGAAAGCACTATATGAAACAGC 1020
1021 CATTTGACTACATGAACTTACCTCTTTATTTAAGACATGAAAGCACTCTCTGTGACTG 1080
1021 CATTTGACTACATGAACTTACCTCTTTATTTAAGACATGAAAGCACTCTCTGTGACTG 1080
1081 AAGTTGCTGAAGATCTCTCAACAAGTTTCAACTGTCCATCTCCAACTGGGCTTACCACTGG 1140
1081 AAGTTGCTGAAGATCTCTCAACAAGTTTCAACTGTCCATCTCCAACTGGGCTTACCACTGG 1140
1141 TTTTATTTGTAGCAACAGGCTACTTATGAACTGATAGAAAGCACTGCGAATGGGTTG 1200
1141 TTTTATTTGTAGCAACAGGCTACTTATGAACTGATAGAAAGCACTGCGAATGGGTTG 1200
1201 CTGAGGCTCTTGGGAAAAAGCAGAGTTCTACTTTGTTAAGGAGACTTTTGGAAAGTGA 1260
1201 CTGAGGCTCTTGGGAAAAAGCAGAGTTCTACTTTGTTAAGGAGACTTTTGGAAAGTGA 1260

1261 ACATTTCTCAAGATGCTAATGTGCTCTCAAAAGCAGAGAGGAGTTCCAGTGGAT 1320
1261 ACATTTCTCAAGATGCTAATGTGCTCTCAAAAGCAGAGAGGAGTTCCAGTGGAT 1320
1321 TTTTGTATTAACATGATGTTGATTTAATATATCTCATGAGGAAAAATATATGACACTG 1380
1321 TTTTGTATTAACATGATGTTGATTTAATATATCTCATGAGGAAAAATATATGACACTG 1380
1381 AGGAAATACAAAGAAATGAGACATGACATGGAAGTCCAGATATATGATGTTCCAGATG 1440
1381 AGGAAATACAAAGAAATGAGACATGACATGGAAGTCCAGATATATGATGTTCCAGATG 1440
1441 ATGAAAGTGCAGAAACTGTTTTCAGAGATAGGAAGAAATTAATCTTTGGAATTCAG 1500
1441 ATGAAAGTGCAGAAACTGTTTTCAGAGATAGGAAGAAATTAATCTTTGGAATTCAG 1500
1501 TGGAACTTAACAGAAAGAACTTAATGCAACAGTGAATGGCTTGCACAGCATGTAATCT 1560
1501 TGGAACTTAACAGAAAGAACTTAATGCAACAGTGAATGGCTTGCACAGCATGTAATCT 1560
1561 TCTATGCTGGTGGCAAGCAGATATCCATGTCATTTTTCGAATCCTATATTTGATGTGCGAG 1620
1561 TCTATGCTGGTGGCAAGCAGATATCCATGTCATTTTTCGAATCCTATATTTGATGTGCGAG 1620
1621 TTTAACTGAAAGCAGACATCTACTATGCTTCTTACTAGAAATTAACCTGCGAGATTTGCT 1680
1621 TTTAACTGAAAGCAGACATCTACTATGCTTCTTACTAGAAATTAACCTGCGAGATTTGCT 1680
1681 ATGTATGTACTAAGCAAAATGTTACTGAATTTCTTCTATCAATAAAGATATACAAAGAGCG 1740
1681 ATGTATGTACTAAGCAAAATGTTACTGAATTTCTTCTATCAATAAAGATATACAAAGAGCG 1740
1741 AGAACCAGTATCTTATGCTGGAATGTTAGAAACCAAGATCTCTAAATTTATCCAGC 1800
1741 AGAACCAGTATCTTATGCTGGAATGTTAGAAACCAAGATCTCTCTAAATTTATCCAGC 1800
1801 TCAACAGAGATTTATATCCAGTGAATATTAACATGATCCAGAAAGCAGAAATATTTAA 1860
1801 TCAACAGAGATTTATATCCAGTGAATATTAACATGATCCAGAAAGCAGAAATATTTAA 1860
1861 GTGGGGAATTAATTAAGACCTCATCTGTATTTCTAGTGTGATGAGTGGGACTATTTA 1920
1861 GTGGGGAATTAATTAAGACCTCATCTGTATTTCTAGTGTGATGAGTGGGACTATTTA 1920
1921 GTCCAACCATGAAGAAAGCAGAAAGAAATTTAGTGAAGCAGGAAACTACTTAAAGAT 1980
1921 GTCCAACCATGAAGAAAGCAGAAAGAAATTTAGTGAAGCAGGAAACTACTTAAAGAT 1980
1981 ATGTATACCTGGAATTTATTTGGAAGAAAGATGTTTGTCTACTGCTCAACCAGATATGCTG 2040
1981 ATGTATACCTGGAATTTATTTGGAAGAAAGATGTTTGTCTACTGCTCAACCAGATATGCTG 2040
2041 CAAAGTCTCCAGCCCTGCTGCTTGCAGACACAGAGGCAAAATTAAGAGAGATCCAC 2100
2041 CAAAGTCTCCAGCCCTGCTGCTTGCAGACACAGAGGCAAAATTAAGAGAGATCCAC 2100
2101 TAGCTAGCACACATGACCAAGACATAGTTCAAAATTAATACAGATGCACTCTGGAATGT 2160
2101 TAGCTAGCACACATGACCAAGACATAGTTCAAAATTAATACAGATGCACTCTGGAATGT 2160
2161 TTTCCGGAATCACTGTGGAAGAAATCTTCCAGTTATTTTCAGACTTCCAGAAACCTTATGGA 2220
2161 TTTCCGGAATCACTGTGGAAGAAATCTTCCAGTTATTTTCAGACTTCCAGAAACCTTATGGA 2220
2221 TTTTGTTCAGTGAATGACGCTGTAATTCCTCAATATTAAGAAAGCAATATGACATGTAAT 2280
2221 TTTTGTTCAGTGAATGACGCTGTAATTCCTCAATATTAAGAAAGCAATATGACATGTAAT 2280
2281 AGCAGAAATCTGATGATCTTACTCATGCTGTTAAATCTAAAGAAATCTCAGATG 2340
2281 AGCAGAAATCTGATGATCTTACTCATGCTGTTAAATCTAAAGAAATCTCAGATG 2340
2341 GGAGAGGAATCTTGGCGGCAATATTTGATCTCTGCTCCCTCTCTCTCTGTTTGG 2400

|||||
Db 2341 GGAGAGAACTTGGGGCATTTTGGATCTCTGCTCCCTCTCTTCTTTGTTTG 2400
OY 2401 TGAATCTGATTCAGTGGCCCAAGATTTTGCATTTCTTCAACACGCGTATATTTGAG 2460
Db 2401 TGAATCTGATTCAGTGGCCCAAGATTTTGCATTTCTTCAACACGCGTATATTTGAG 2460
OY 2461 AAAACCTTGTATTTGGGCGAAGAAATTTAGACGAGACTAGAAAATCTATCACAAAT 2520
Db 2461 AAAACCTTGTATTTGGGCGAAGAAATTTAGACGAGACTAGAAAATCTATCACAAAT 2520
OY 2521 TACCTGCTCAAGATGAAACCTCTCTTCCAGCTTATGATTTTCTAAGTATGATAGATG 2580
Db 2521 TACCTGCTCAAGATGAAACCTCTCTTCCAGCTTATGATTTTCTAAGTATGATAGATG 2580
OY 2581 CCGCAACATCTCAAGCTGCACTAGGAAAGTCCCAAGTGTATGAAAGAAACAGATGTGC 2640
Db 2581 CCGCAACATCTCAAGCTGCACTAGGAAAGTCCCAAGTGTATGAAAGAAACAGATGTGC 2640
OY 2641 AGGAAATGATTAAGGAAACATGAAAGATTAATCGGACGTGAGAAAGAACCGATTGAA 2700
Db 2641 AGGAAATGATTAAGGAAACATGAAAGATTAATCGGACGTGAGAAAGAACCGATTGAA 2700
OY 2701 CTCTGAGAAATAAGCATTGGAAATAGAAATTTGTTTAAAGACGAGAAAATCATTTA 2760
Db 2701 CTCTGAGAAATAAGCATTGGAAATAGAAATTTGTTTAAAGACGAGAAAATCATTTA 2760
OY 2761 GACGATTAAGAGTTAGATGCTCAAAAGTAACTAATTTATAGGCTGTGGTTTCCA 2820
Db 2761 GACGATTAAGAGTTAGATGCTCAAAAGTAACTAATTTATAGGCTGTGGTTTCCA 2820
OY 2821 AAATTTTGGCATGATAGACTAATTTATTTCTTAAAGATATATTAATATCATTTTC 2880
Db 2821 AAATTTTGGCATGATAGACTAATTTATTTCTTAAAGATATATTAATATCATTTTC 2880
OY 2881 AAGTTTGCAGACTAGTGGCATCAATAGATTAATATTAATTAATTAATTAATTA 2940
Db 2881 AAGTTTGCAGACTAGTGGCATCAATAGATTAATATTAATTAATTAATTAATTA 2940
OY 2941 ATTTCTGTACTACTAATTAACAAAGTAAAGTGAGGCGCAAAATTAATTTGTATTT 3000
Db 2941 ATTTCTGTACTACTAATTAACAAAGTAAAGTGAGGCGCAAAATTAATTTGTATTT 3000
OY 3001 ACTTTTCAACCCAGTAGATATACCCAAATATGCGAAATATAGAAATTTAATGAGATTT 3060
Db 3001 ACTTTTCAACCCAGTAGATATACCCAAATATGCGAAATATAGAAATTTAATGAGATTT 3060
OY 3061 TACATCTTTTGTACCAAGCTTCTAATATGACATATTTTATCTACTGCAATTT 3120
Db 3061 TACATCTTTTGTACCAAGCTTCTAATATGACATATTTTATCTACTGCAATTT 3120
OY 3121 CTATCTCGAGTACCAATTTTCAAGTGTGATGCGCATGCGGCGGTGACATCTT 3180
Db 3121 CTATCTCGAGTACCAATTTTCAAGTGTGATGCGCATGCGGCGGTGACATCTT 3180
OY 3181 ATTGACAGTTCAGTACTAGCAAAACCTAGCATTAATTAATTTAGTTCAGGCATGATTT 3240
Db 3181 ATTGACAGTTCAGTACTAGCAAAACCTAGCATTAATTAATTTAGTTCAGGCATGATTT 3240
OY 3241 CTATTTGGATTTAAATTAACCTTAATCAAGTAACTCCACAGTGCATTTCAAGCTG 3300
Db 3241 CTATTTGGATTTAAATTAACCTTAATCAAGTAACTCCACAGTGCATTTCAAGCTG 3300
OY 3301 ACAGTATATTTGTTTATTTGAGTCATGATTAATAATCAGCGTTTGCACCTCAGG 3360
Db 3301 ACAGTATATTTGTTTATTTGAGTCATGATTAATAATCAGCGTTTGCACCTCAGG 3360
OY 3361 GATATTTAGCAATTTGCGGAGACATTTTGTGATGATGATGAGGCACTTTTGCATTT 3420
Db 3361 GATATTTAGCAATTTGCGGAGACATTTTGTGATGATGATGAGGCACTTTTGCATTT 3420
OY 3421 TAGTAGTAGAGGCGCATGATCTGCTAATTAATCTGCACTTGGACAGCGCCACACAA 3480
|||||

Db 3421 TAGTAGTAGAGGCGCATGATCTGCTAATAATCACTGATTTGGACAGCCGCCACACAA 3480
OY 3481 AGAATTAATCTGCGCGAATATGATGTCGCGCAAGGCTAGTAACCTTGTTTAAAGTA 3540
Db 3481 AGAATTAATCTGCGCGAATATGATGTCGCGCAAGGCTAGTAACCTTGTTTAAAGTA 3540
OY 3541 ACCGTGGCAGACTAGTTCAGAAATTTCTGTTTCTCTCACTATGATGTTTGA 3600
Db 3541 ACCGTGGCAGACTAGTTCAGAAATTTCTGTTTCTCTCACTATGATGTTTGA 3600
OY 3601 AATTTTGGCTATTTAAAGATATGATTAATGATGCTTATCTGATTTATTAACCTGATACAA 3660
Db 3601 AATTTTGGCTATTTAAAGATATGATTAATGATGCTTATCTGATTTATTAACCTGATACAA 3660
OY 3661 CTGTGATCTTCTAATTTTCAAGAGATGAGGGAATACCCAGAGAGAGACTCGAAT 3720
Db 3661 CTGTGATCTTCTAATTTTCAAGAGATGAGGGAATACCCAGAGAGAGACTCGAAT 3720
OY 3721 GATATTTATATTTTAAAGTATGATTTAAACCTCTCTATTTCTACAGTTATATGCTA 3780
Db 3721 GATATTTATATTTTAAAGTATGATTTAAACCTCTCTATTTCTACAGTTATATGCTA 3780
OY 3781 AATTTCAATGTAACAGGATTCAGCATTTGCCATCTCTCTCATGGAAGAGAGCTCCC 3840
Db 3781 AATTTCAATGTAACAGGATTCAGCATTTGCCATCTCTCTCATGGAAGAGAGCTCCC 3840
OY 3841 TCATCTGAAGCGCTCTGAAATCTACCTTGAAGTCTGAGCAATTCAGTATGATCTCC 3900
Db 3841 TCATCTGAAGCGCTCTGAAATCTACCTTGAAGTCTGAGCAATTCAGTATGATCTCC 3900
OY 3901 TGAGCCACAGCGCTCATCTGTGAGGAGGAGAAATTAAGCAAGATTAATTTTCA 3960
Db 3901 TGAGCCACAGCGCTCATCTGTGAGGAGGAGAAATTAAGCAAGATTAATTTTCA 3960
OY 3961 TCCAAATCTATTAAGCTGTTAGACTGATCTGTTTGTAGCACTTGTGCTCATTTTGGCT 4020
Db 3961 TCCAAATCTATTAAGCTGTTAGACTGATCTGTTTGTAGCACTTGTGCTCATTTTGGCT 4020
OY 4021 CTGTGATTTTGTGAGACATTTGTTGAGATATTTATTTGGGCTCTGATATTTTTC 4080
Db 4021 CTGTGATTTTGTGAGACATTTGTTGAGATATTTATTTGGGCTCTGATATTTTTC 4080
OY 4081 TTTTAAATCTACTGATATCTTGTCTTAAATTTCTTCAATATGATTTTGCCTGAT 4140
Db 4081 TTTTAAATCTACTGATATCTTGTCTTAAATTTCTTCAATATGATTTTGCCTGAT 4140
OY 4141 ACAACGATTTTATTAAGCAATTTAAGGAATCTAAGCTTAAACCTAGTAAGTCA 4200
Db 4141 ACAACGATTTTATTAAGCAATTTAAGGAATCTAAGCTTAAACCTAGTAAGTCA 4200
OY 4201 MTATTTCTTATTAACATAGACCGGCTGTAACCTGACACCCCTGCAATTTTTC 4260
Db 4201 MTATTTCTTATTAACATAGACCGGCTGTAACCTGACACCCCTGCAATTTTTC 4260
OY 4261 CTGTAGCATGATGAGCTGATTTAAACCTCAATTTCTTATTTCAATATGAGAA 4320
Db 4261 CTGTAGCATGATGAGCTGATTTAAACCTCAATTTCTTATTTCAATATGAGAA 4320
OY 4321 CAATGAGATGAACCTTAATATAGTGTAGTAATAAATCATTAATTAATTA 4380
Db 4321 CAATGAGATGAACCTTAATATAGTGTAGTAATAAATCATTAATTAATTA 4380
OY 4381 GAAATGCTAATTAATAGTACAGACATTAAGCAATTAATGCTTGTGATGATGATCTT 4440
Db 4381 GAAATGCTAATTAATAGTACAGACATTAAGCAATTAATGCTTGTGATGATGATCTT 4440
OY 4441 GTGAGCAATTTTACAGCTATTAATGTTTGTCAATAATTTTAAATAAAGTGTGAGGTTTC 4500
Db 4441 GTGAGCAATTTTACAGCTATTAATGTTTGTCAATAATTTTAAATAAAGTGTGAGGTTTC 4500
OY 4501 AGAATACCTTCAAAAAA 4526
Db 4501 AGAATACCTTCAAAAAA 4526

```
RESULT 2
US-09-764-864-358
; Sequence 358, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 358
; LENGTH: 1423
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (32)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1369)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-358

Query Match          2.3%; Score 106; DB 10; Length 1423;
Best Local Similarity 63.2%; Pred. No. 1.4e-13;
Matches 230; Conservative 0; Mismatches 125; Indels 9; Gaps 4;

QY 2844 TAATTTATTTCTTAAAGAAATATATTAATTCATTTTCAGTTTGCAGACTAGTCCATCC 2903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 651 TAAACTTTGGCTACTCTCATTAATTAACCACTTATATAGTTCTAGAGCAGCGCTGTCC 710

QY 2904 AATAGAAATTATATATATAGTACATATTTTAAATTTTCTAGTACTACTTAAAC 2963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 711 AATAGAAATTATATATAGTACATATTTTAAATTTTCTAGTACTACTTAAAC 768

QY 2964 AAGTAAAGTGAAGCAGGCAAAA-----TAATTTGATATTTACTTTTCAACCCAGTAGTA 3018
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 769 AAGTAAAGTGAAGCAGGCAAAA-----TAATTTGATATTTACTTTTCAACCCAGTAGTA 828

QY 3019 TACCCAAATAGCGAAATATATAGAAATTTATATGAGATATTTTACATCTTTT-TTGTAC 3077
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 829 ATTGAACATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 888

QY 3078 CAAGTCTTCTAATATGCGATCATATTTTATCTACTGCAATTTTCTTCCGAGTAGCC 3137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 889 CTAGTCTTCTAATATGCGATCATATTTTATCTACTGCAATTTTCTTCCGAGTAGCC 948

QY 3138 ATATTCAGATGTTCA-TTGCACATGTGGCGCTGACTACTGTATGAGCAGACTAGTA 3196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 949 ACATTGCAAGTGTCTCAGTAGCCACATGTGGCTAGTGTCTACTGACGTGACAGCAGACTT 1008

QY 3197 CTAG 3200
    ||||
DB 1009 CTAG 1012
```

```
RESULT 3
US-10-121-235-20
; Sequence 20, Application US/10121235
; Publication No. US20030032609A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-Bn

; APPLICANT: Maemura, Koji
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: METHODS OF MODULATING OF ANGIOGENESIS
; FILE REFERENCE: 05433/037001
; CURRENT APPLICATION NUMBER: US/10/121,235
; CURRENT FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: US 09/374,454
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: US 60/096,515
; PRIOR FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 6792
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-121-235-20

Query Match          2.2%; Score 97.8; DB 9; Length 6792;
Best Local Similarity 65.2%; Pred. No. 1.9e-11;
Matches 208; Conservative 0; Mismatches 102; Indels 9; Gaps 4;

QY 2889 AGACTAGTCCATCCATAGCATATATATATATATCTACATATTTTATTAATTTCTA 2948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4797 AGAGCAGCGCTGTCCATAGAAATATATCTGAGCCACATGTATATTT--TATTTCTT 4854

QY 2949 GTACTACATTAAACAAGTAAAGTAGCAGGCAAAA-----TAATTTGATATTTACT 3003
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4855 CTAGCCACATTAAAGAGTAAAGATACAGATGAAGCTATATTTATGTTTATTCAG 4914

QY 3004 TTTCACCCAGTAGTATACCCAAATAGCGAAATATATAGAAATTTATATGAGATTTTAC 3063
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4915 TATATCCAAATATCATTTTGACATGTAAATTAATTAATTAATTAATTAATTAATTAATTA 4974

QY 3064 ATCTTTT-TTGTACCAAGCTTCTTAATGACATATTTTACTTCTGCAATTTCT 3122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4975 ATCTTTTGTATATAGTATCTTAATATCTGATATCTTATCTTACATTTGATGACATCT 5034

QY 3123 TACTTCCAGTAGCCATATTTCAAGTGTCA-TTGCACATGTGGCGCTGACTACTGTA 3181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5035 CACTTTGTACTAGCCACATTTGCAAGTGTCTGATGACCATATGTGGCTAGTACTGCA 5094

QY 3182 TTGCACAGTTCACTAGTAG 3200
    ||||| ||| ||| |||
DB 5095 CTGACAGCAGCAGTTCTAG 5113

RESULT 4
US-09-852-067-3/C
; Sequence 3, Application US/09852067
; Patent No. US20020076777A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; FILE REFERENCE: CL000897-CIP
; CURRENT APPLICATION NUMBER: US/09/852,067
; CURRENT FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 31208
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(31208)
; OTHER INFORMATION: n = A,T,C or G
US-09-852-067-3

Query Match          1.9%; Score 86.8; DB 10; Length 31208;
```

Best Local Similarity	59.3%;	Pred. No. 1e-08;
Matches 185;	Conservative 0;	Mismatches 122;
	Indels 5;	Gaps 2

OY	2892	CTAGAGCCATCCCATAGATTTATTAATATAGTCACCATATTTTATTTAAATTTTCTAGTA	2951
Db	27847	CTAGCACCAGCATCAATATGAAATATATATGTGATCTACATGTGTATTTAAAAATGTTCTGGTG	27788
OY	2952	ACTACATTTAAACAAGTAAAGTGAGCAGCGCAAAATTAATTTGATATTAATCTTTACACC	3011
Db	27787	GTCACATTTTTTAAAGGTGAAGTTAAT---CTTAATAGAGTTATTTAAACCAAAATTA	27732
OY	3012	ACTGATATCCCAAAATATAGCAAAATATAGAAATTAATATAGATATTTATACATCCCTTT	3071
Db	27731	CCATTCACACCGTATCATCATATTAATTAATTAATTAATTAACAGAAATGTTTCACATTTTTTTA	27672
OY	3072	TTGTACCAAGCTCTTCAAAATGCAAGTACATATTTTATTAATCTTACTGCATTTTCTTACTTCGA	3131
Db	27671	GGGTACTAATGTTTCAAAATCCAGCACGATATTTTAAACTTACGCAACATCTCAATTCOA	27612
OY	3132	GTAGCCATATTTCAAGTGTCCA-TTGCCACATGTGGCGCTGTGACTCTGATATGGACAGT	3190
Db	27611	CTGGGCACATGTCAATGTGCTCAGTAGCCGTATGTGGCCCGGTGCTACTACACGCGGCAGT	27552
OY	3191	TCAGTACTAGAC 3202	
Db	27551	GTACCTTCAGAC 27540	

RESULT 5
US-09-901-136-3/c

```

? Sequence 3, Application US/09901136
? Publication No. US20030039968A1
?
? GENERAL INFORMATION:
?
? APPLICANT: HU, Song et al.
?
? TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
? TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
? TITLE OF INVENTION: USES THEREOF
? FILE REFERENCE: CLO01273
? CURRENT APPLICATION NUMBER: US/09/901,136
? CURRENT FILING DATE: 2001-07-10
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: FastSeq for Windows Version 4.0
?
? SEQ ID NO 3
?
? LENGTH: 378361
?
? TYPE: DNA
?
? ORGANISM: Human
?
? FEATURE:
?
? NAME/KEY: misc_feature
? LOCATION: (1)..(378361)
? OTHER INFORMATION: n = A,T,C or G
?
? US-09-901-136-3

```

Query Match	1.8%	Score 82.4;	DB 9;	Length 378361;
Best Local Similarity	57.9%	Pred. No. 3.1e-07;		
Matches 183; Conservative	0;	Mismatches 131;	Indels 2;	Gaps 2

OY	2889	AGACGAGGCCATCCAAATAGAAATTATAATATAGTCACATATTTTATTTAAATTTTCTA	2348
Db	239445	AATACGACCTCTTCAACGAATATATATGTGCTATATATTAATTTTCTAA	239386
OY	2949	GTAACCTACATTAACCAAACTAAAGTAGCAGGCGCAAAATTAATTGATTTACTTTTCA	3008
Db	239385	TAGTCTCATTAATAAATTGTACGAGATTAATTTTAAATAGATATTTTATTCGACCATA	239326
OY	3009	C-CGCGTAGTATATACCAAAATAGCGAATATAGAAATTATTAATGAGATATTTTACATCC	3067
Db	239325	CATCTAAATATATATTTCAACATATATCAATATATAAAATTAATGAGATTTTAAATCT	239266
OY	3068	TTTTTTGTACGAAGTCTTAAATGACATACATATTTTAACTACGCACTTCTACTT	3127
Db	239265	TTTTTCATCCCAAGTCTTAAATTTCTGGTTTACATTTTAAACCACTATTAATT	239206
OY	3128	CCGAGTAGGCATATTTCAAGGTCTTCATTCGCCACATGTGGCTGTGACTACTGATTTGGAC	3187

Db 239205 CAGTTGACACAAATTTCAATGCTGATAGC-CATGTTTCCAGTACTAGCCTATTTAGC 239147
 Oy 3188 AGTCAGTACTAGACA 3203
 Db 239146 GGCACAGGCTTAGACA 239131

RESULT 6
US-09-731-872-239/c
; Sequence 239, Application US/09731872

```

: Patent NO. US20020102604A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Dumas Milne Edwards, Jean Baptiste
: APPLICANT: Bouguetieret, Lydie
: APPLICANT: Jobert, Severin
: TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
:
: FILE REFERENCE: 78 US3.REG
:
: CURRENT APPLICATION NUMBER: US/09/731,872
:
: CURRENT FILING DATE: 2000-12-07
:
: PRIOR APPLICATION NUMBER: US 60/169,629
:
: PRIOR FILING DATE: 1999-12-08
:
: PRIOR APPLICATION NUMBER: US 60/187,470
:
: PRIOR FILING DATE: 2000-03-06
:
: NUMBER OF SEQ ID NOS: 482
:
: SOFTWARE: Patent.pm
:
: SEQ ID NO 239
:
: LENGTH: 1083
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: 421..768
:
: US-09-731-872-239

```

Query Match	1.8;	Score	81.4;	DB	10;	Length	1083;
Best Local Similarity	68.68;	Pred.	No. 3.2e-08;				
Matches 177; Conservative	0;	Mismatches	56;			Indels	2;
						Gaps	1;

QY	3024	AAATAGCGAAATATACAAATTTATATAGAGATATTTTACATCCCTTTT	--GTACCAAG	3081
Db	1014	AATAAATGTAGAAAAAAATTTGTTAATGAAATTTTACATTCCTTTTGTGACTAAG	955	
QY	3082	TCTCTAAATGCAGTACATATATTTTATACAGCATTTCTTACTTCCGAGTAGCCATAT	3141	
Db	954	TCTTGGAATGTGGTGTCTTCTCTTATTTACAGCAATCTCAATGTGGGTAGGCCAAT	895	
QY	3142	TTCAAGTGTTCATTTGCCACATGTGGCCCTGTGACTACTGATTTAGACAGTTCACTAGTA	3201	
Db	894	TTCAAGTGTCTCAGTAGTGCAATGTGGCTAGTGGCCGCAATGTTGGAGCGACAGTCTATA	835	
QY	3202	CAAA	3206	
Db	834	CTATA	830	

```

; RESULT 7
; US-09-864-761-11995
; Sequence 11995, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456

```


3038 TAGAATATTAAAGAGATATTTTACATCCTTTTGTACCAAGTCTCTAAATGCAGTA 3097

```

; APPLICANT: Barrington-Martin, Rory
; APPLICANT: Parker, Alexander
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING

```

; APPLICANT: Parker, Alexander
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING


```

P100001      PRIOR FILLING DATE: 2000-04-11
P100001      PRIOR APPLICATION NUMBER: US 60/196,528
P100001      PRIOR FILLING DATE: 2000-04-11
P100001      PRIOR APPLICATION NUMBER: US 09/687,837
P100001      PRIOR FILLING DATE: 2000-10-13
P100001      PRIOR APPLICATION NUMBER: US 60/240,503
P100001      PRIOR FILLING DATE: 2000-10-13
P100001      PRIOR APPLICATION NUMBER: US 60/240,508
P100001      PRIOR FILLING DATE: 2000-10-13
P100001      PRIOR APPLICATION NUMBER: US 60/240,539
P100001      PRIOR FILLING DATE: 2000-10-13
P100001      PRIOR APPLICATION NUMBER: US 60/240,543
P100001      PRIOR FILLING DATE: 2000-10-13
P100001      NUMBER OF SEQ ID NOS: 134
P100001      SOFTWARE: PatentIn Ver. 2.1
P100001      SEQ ID NO 86
P100001      LENGTH: 66686
P100001      TYPE: DNA
P100001      ORGANISM: Homo sapiens
P100001      FEATURE:
P100001      OTHER INFORMATION: ordered human genomic DNA at CLAF-5 locus
US-09-736-960-86

Query Match          . 1.6%; Score 72.2; DB 10; Length 66686;
Best Local Similarity 59.7%; Pred. No. 2.4e-05;
Matches 181; Conservative 0; Mismatches 108; Indels 14; Gaps 3;

QY 2913 ATATATTAAGTCACATATTTTATTTTAAATTTTCTAGTAATCTATTAACCAAGTAAAGTAA 2972
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 24060 ACAATGCGCGCCACATATGTCATTTTATATTTGTGTGCTAACCAATTAAGTAAAGA 24001

QY 2973 GTGACGAGGGCAAAATATTTTGTATTTACTTTTCCACCAGTGTATACCAAAATATAGC- 3031
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 24000 AAAAGTGAATATTTATTTTAAATATATTTTGTGTTTATCTTAAGATTAAGTAAATATCTGCTCC 23941

QY 3032 -----GAATATATGAATATTTATTAATGAGATATTTTACATCTTTTGTGACCAAGTCTT 3085
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 23940 ATGTAAACAAATATTAATAATATACAGTAGTAATATTTATCTTTTATTCACAAATAGGCTT 23881

QY 3086 CTAAATGACAGTACATATTTTAACTT-----ACTGCATTTCTTACTTCGAGTAGCCAT 3139
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 23880 CAAATATGTGGCTATTTTTCACCTTATCTTGATTTATATTCACAAATTTGAAGTACCAT 23821

QY 3140 ATTTCGANG-TGTTGATTCGCCACATGTGGCGTGACTACGTATTTGGACAGTTCAGTACT 3198
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 23820 ATTTCAAGCAATTGATATGTCACATGTGCTAATAGGCTACTGTAATGGACAGCACAATCT 23761

QY 3199 AGA 3201
      |||
Db 23760 AGA 23758

RESULT 14
US-09-822-246-3/c
Sequence 3, Application US/09822246
Patent No. US20020142383A1
GENERAL INFORMATION:
APPLICANT: MERRILOV, Gennady et al.
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001149
CURRENT APPLICATION NUMBER: US/09/822,246
CURRENT FILING DATE: 2001-04-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 197997
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(197997)

```



```

Db 1527 GAAATTGGGCAATATTTTACACTTAVGGACATCCATTCATTCAGACATCCACATTTCA 1586
QY 3147 GTGTTTC-TTGCACATATGGCCCTGGATACGTATGGACATGTCAGTACTA 3199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1587 GTGTCACGTGGCTGCATGTGTCCCTGGTGGCTACCATATTTGGACAGCACAGGTCTA 1640

Search completed: March 22, 2003, 14:37:45
Job time : 2155 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2003, 08:55:02 ; Search time 1200 Seconds
(without alignments)
9783.383 Million cell updates/sec

Title: US-09-847-046-1
Perfect score: 4526
Sequence: 1 gccgcagtcgcygcygctgc.....ccttcaaaaaaaaaaaaaa 4526

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4747046 seqs, 1296957505 residues

Total number of hits satisfying chosen parameters: 9494092

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : Pending_Patents_MA_New.*
1: /cgn2_6/ptodata/2/pna/PCF_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4016.6	88.7	4135	US-10-342-887-77	Sequence 77, Appl
2	934.6	20.6	3756	US-10-144-771-15372	Sequence 15372, A
3	354.2	7.8	435	US-09-912-293-92043	Sequence 92043, A
4	341	7.5	335	US-09-513-999C-32911	Sequence 32911, A
5	312	6.9	313	US-09-912-293-91554	Sequence 91554, A
6	265.8	5.9	496	US-09-912-293-96911	Sequence 96911, A
7	257	5.7	257	US-10-349-781-15516	Sequence 15516, A
8	245	5.4	245	US-10-349-781-36747	Sequence 36747, A
9	228	5.0	228	US-10-349-781-24291	Sequence 24291, A
10	188.6	4.2	375	US-09-912-293-18131	Sequence 18131, A
11	181	4.0	181	US-09-539-806B-30366	Sequence 30366, A
12	179	4.0	191	US-10-349-781-1729	Sequence 1729, Ap
13	171	3.8	176	US-10-349-781-23200	Sequence 23200, A
14	108.4	2.4	1043426	US-09-947-911-240	Sequence 240, App
15	103.2	2.3	76321	US-09-949-002-578	Sequence 578, App
16	103.2	2.3	76368	US-09-949-002-803	Sequence 803, App
17	102.8	2.3	2267436	US-09-947-911-230	Sequence 230, App
18	98.8	2.2	2682138	US-09-947-911-255	Sequence 255, App
19	98.8	2.2	195909	US-09-949-004-469	Sequence 469, App
20	98.2	2.2	195910	US-09-949-004-682	Sequence 682, App
21	97.8	2.2	6792	US-10-126-052A-165	Sequence 165, Appl
22	93.4	2.1	302603	PCF-US02-32700-8	Sequence 8, Appl1
23	93.4	2.1	302603	US-10-271-416-8	Sequence 8, Appl1
24	89.4	2.0	495619	US-09-947-911-55	Sequence 55, Appl

25	87	1.9	96596	1	PCF-US02-41414-1022	Sequence 1022, Ap
26	86.8	1.9	31208	8	US-10-338-691-3	Sequence 3, Appl1
27	86.2	1.9	66380	5	US-09-947-911-327	Sequence 327, Appl
28	85	1.9	2593930	5	US-09-947-911-54	Sequence 54, Appl
29	84.2	1.9	1059516	5	US-09-947-911-86	Sequence 86, Appl
30	83.6	1.8	481258	5	US-09-947-911-94	Sequence 94, Appl
31	83.2	1.8	2865598	5	US-09-947-911-26	Sequence 26, Appl
32	82.4	1.8	13712	8	US-10-311-455-1504	Sequence 1504, Ap
33	81.2	1.8	366	6	US-09-912-293-117056	Sequence 117056, Ap
34	80.6	1.8	601	6	US-09-949-004-6559	Sequence 6559, Ap
35	79.6	1.8	145606	9	US-60-444-614-7	Sequence 7, Appl1
36	79.4	1.8	601	6	US-09-949-004-6560	Sequence 6560, Ap
37	79.4	1.8	71848	6	US-09-949-004-691	Sequence 691, App
38	79.4	1.8	72178	6	US-09-949-004-471	Sequence 471, App
39	79.4	1.8	3436920	5	US-09-947-911-49	Sequence 49, Appl
40	79.2	1.7	708988	5	US-09-947-911-172	Sequence 172, App
41	79	1.7	59998	6	US-09-949-004-608	Sequence 608, App
42	78.8	1.7	32433	1	PCF-US02-41414-1148	Sequence 1148, Ap
43	78.4	1.7	488	6	US-09-912-293-13634	Sequence 13634, A
44	77.8	1.7	151152	9	US-60-446-133-243	Sequence 243, App
45	77.8	1.7	2605602	5	US-09-947-911-314	Sequence 314, App

ALIGNMENTS

```

RESULT 1
US-10-342-887-77
; Sequence 77, Application US/10342887
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 77
; LENGTH: 4135
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-77

Query Match      88.7% Score 4016.6; DB 8; Length 4135;
Best Local Similarity 98.7%; Pred. No. 0; Mismatches 0; Indels 54; Gaps 1;
Matches 4080; Conservative 1;

QY 431 TTTTACATGCCAAGTAAGTCTTTTACAGAGAGAGTCTCAGAAATATTTTGTAC 490
    |||||||
DB 1 TTTTTCATGCCAAGTAAGTCTTTTACAGAGAGAGTCTCAGAAATATTTTGTAC 60

QY 491 ATTGCACAGC-----
    |||||||
DB 61 ATTGCACAGCAGAAAGCCTTTTACGATTTTGTCAAGCTGATTCCTCCAGAACATC 120

QY 503 -----TCTTGAAGACTGAATGAGCTGTGACCTGCGAGACTATGAAATTTCAGT 556
    |||||||
DB 121 TGTATTCTTGAAGACTGAATGAGCTGTGACCTGCGAGACTATGAAATTTCAGT 180

QY 557 TGCAGAGTTAATTTGCTGCAAGAAAGAAATATCAAGATCTGTGAAAGAAAGATTTT 616
    |||||||

```

Dh 181 TGCCAGGTTAATTGTGTCTCAAGAAAGAAATATACAGATCTGTGGAAAAAGAAAGATTT 240
Qy 617 GATGAAGCATATTTATTTCAAGGGCAACATATTTGCTCAGAGAAATTCCTTACTGACACCTT 676
Dh 241 GATGAAGCATATTTATTTCAAGGGCAACATATTTGCTCAGAGAAATTCCTTACTGACACCTT 300
Qy 677 GTTTGATGTGAATGCCATTTGTGCCCATGTTCTCTTGTCTCTTTTAACTGAAGTGA 736
Dh 301 GTTTGATGTGAATGCCATTTGTGCCCATGTTCTCTTGTCTCTTTTAACTGAAGTGA 360
Qy 737 ATATATATACCAACCTGGAAGCCTTCAGAACATAGAAAATTCCTGGAAGGAAAAAGCAA 796
Dh 361 ATATATATACCAACCTGGAAGCCTTCAGAACATAGAAAATTCCTGGAAGGAAAAAGCAA 420
Qy 797 TATATATATTCATATGTGAAGGCCATTGGAATACAGAGACAGAGCATGATGAAGC 856
Dh 421 TATATATATTCATATGTGAAGGCCATTGGAATACAGAGACAGAGCATGATGAAGC 480
Qy 857 CGGTTTGTGTATGGGACTACATACCAATTTGTCTTAACCAAGAAATTCCTTTTGA 916
Dh 481 CGGTTTGTGTATGGGACTACATACCAATTTGTCTTAACCAAGAAATTCCTTTTGA 540
Qy 917 AAGTATGGGCTCTGAGAGTGAATATGCACATCTCTACTTTTTCATTTGAATAGT 976
Dh 541 AAGTATGGGCTCTGAGAGTGAATATGCACATCTCTACTTTTTCATTTGAATAGT 600
Qy 977 CTGTGACTTGACCCAGCAATGTAGAGAACATATGGAACGCCATTGACTACACTGAA 1036
Dh 601 CTGTGACTTGACCCAGCAATGTAGAGAACATATGGAACGCCATTGACTACACTGAA 660
Qy 1037 CATTCACCTGTTTATTAAGCAATGAAGCAACCTCTGTGACTGAAGTGTGAAGATCC 1096
Dh 661 CATTCACCTGTTTATTAAGCAATGAAGCAACCTCTGTGACTGAAGTGTGAAGATCC 720
Qy 1097 TCAACAAGTTTCAACTCTCCATCTCAACTGGGCTTACCACTGGTTTATTTAGTGAACA 1156
Dh 721 TCAACAAGTTTCAACTCTCCATCTCAACTGGGCTTACCACTGGTTTATTTAGTGAACA 780
Qy 1157 ACAGGCTACTTATGAGCTGTATAGAGAACTGCAAAATGGTGTCTTGGGCTTCTGGG 1216
Dh 781 ACAGGCTACTTATGAGCTGTATAGAGAACTGCAAAATGGTGTCTTGGGCTTCTGGG 840
Qy 1217 AAAAGGAGGAGTCTACTCTTGTAAAGGAGCTCTTGTGAAGTGAACATTCCTCAAGATGC 1276
Dh 841 AAAAGGAGGAGTCTACTCTTGTAAAGGAGCTCTTGTGAAGTGAACATTCCTCAAGATGC 900
Qy 1277 TAAATGCTCTTCAAAAAGAGAGAGAGAGGAGTCCAGTGGAAATTTTGTGATTAACATGA 1336
Dh 901 TAAATGCTCTTCAAAAAGAGAGAGAGAGGAGTCCAGTGGAAATTTTGTGATTAACATGA 960
Qy 1337 TGTGTGATTTAATATATCTCATGTGGAATAATATATGCACATTTGAGGAATAATACAGAGA 1396
Dh 961 TGTGTGATTTAATATATCTCATGTGGAATAATATATGCACATTTGAGGAATAATACAGAGA 1020
Qy 1397 TGAAGACAATGACATGGAAGGTCAGATATAGATGTCAGAGATGATGAAGTGGAGAAAC 1456
Dh 1021 TGAAGACAATGACATGGAAGGTCAGATATAGATGTCAGAGATGATGAAGTGGAGAAAC 1080
Qy 1457 TGTTTTCAGAGATAGGAAGAAAATTTACCTTTGGAACCTTACAGTGGAACTACAGAGA 1516
Dh 1081 TGTTTTCAGAGATAGGAAGAAAATTTACCTTTGGAACCTTACAGTGGAACTACAGAGA 1140
Qy 1517 AACATTTAATGCAACAGTATGCTTGTGACAGCATATGATCTTCTATGCTGTGGCA 1576
Dh 1141 AACATTTAATGCAACAGTATGCTTGTGACAGCATATGATCTTCTATGCTGTGGCA 1200
Qy 1577 AGCAGTATCATGSCATTTTTCGAATCTATATGATGTGGCACTTAACTGAAGGCAAC 1636
Dh 1201 AGCAGTATCATGSCATTTTTCGAATCTATATGATGTGGCACTTAACTGAAGGCAAC 1260
Qy 1637 ATCTACATGCTTCTTACTAGAAATTAACCTGAGATGTGGTCTGATATGATTAAGCA 1696
Dh 1261 ATCTACATGCTTCTTACTAGAAATTAACCTGAGATGTGGTCTGATATGATTAAGCA 1320

Qy 1697 AAATGTTACTGAATTTTCTTATCATAAAGATGTACAGAAAGCGAGAACCCAGTATCTTA 1756
Dh 1321 AAATGTTACTGAATTTTCTTATCATAAAGATGTACAGAAAGCGAGAACCCAGTATCTTA 1380
Qy 1757 TGTGTGAATGTTAGGAACCAAGATCTCTTAAATTTATCCAGCTCAACAGGATTTCTATA 1816
Dh 1381 TGTGTGAATGTTAGGAACCAAGATCTCTTAAATTTATCCAGCTCAACAGGATTTCTATA 1440
Qy 1817 TTCAGTGAATTAACATGATCCAGAAAGCGAAGAAATATTAAGTGGGGAATTAATATA 1876
Dh 1441 TTCAGTGAATTAACATGATCCAGAAAGCGAAGAAATATTAAGTGGGGAATTAATATA 1500
Qy 1877 AGACCTATCTTGTATTTCTAGTGTGATGATGATTTGGAGCTATTTAGTCCAAACATGAAAAC 1936
Dh 1501 AGACCTATCTTGTATTTCTAGTGTGATGATGATTTGGAGCTATTTAGTCCAAACATGAAAAC 1560
Qy 1937 AGCAAAAAGAAATTTTATGTAAGCAGGAACCTACTTAAAGATATGTATCATCTGGAAAT 1996
Dh 1561 AGCAAAAAGAAATTTTATGTAAGCAGGAACCTACTTAAAGATATGTATCATCTGGAAAT 1620
Qy 1997 TTATTTGGAAGAAATGTTTGTACTGTCAACCAATATGCTGCAAGTCTTCCAGCCCT 2056
Dh 1621 TTATTTGGAAGAAATGTTTGTACTGTCAACCAATATGCTGCAAGTCTTCCAGCCCT 1680
Qy 2057 GCTGCTTGCCAGACACAGAGGCAAAATATGAGAGCATCCCACTAGCTACACATGC 2116
Dh 1681 GCTGCTTGCCAGACACAGAGGCAAAATATGAGAGCATCCCACTAGCTACACATGC 1740
Qy 2117 ACAAGACATGTTCAAAATTAACAGATGCCTCTGGAATTTTCCGGAATCACCTGT 2176
Dh 1741 ACAAGACATGTTCAAAATTAACAGATGCCTCTGGAATTTTCCGGAATCACCTGT 1800
Qy 2177 GGAATAATCTTCCAGTTATTTTCAAGCTTCAGAAACCAATATTTGTTTGTCTCACTGATGG 2236
Dh 1801 GGAATAATCTTCCAGTTATTTTCAAGCTTCAGAAACCAATATTTGTTTGTCTCACTGATGG 1860
Qy 2237 CACTGTAAATCTCAATATTAAGAAAGCAATATTTGACACCTGTTAAGCAAGAAATCTTGGGA 2296
Dh 1861 CACTGTAAATCTCAATATTAAGAAAGCAATATTTGACACCTGTTAAGCAAGAAATCTTGGGA 1920
Qy 2297 TTTCAATTTACTGATGCTGTTAAATCTAAGAAATGTCAGTGGGAGAGAAATCTTGGG 2356
Dh 1921 TTTCAATTTACTGATGCTGTTAAATCTAAGAAATGTCAGTGGGAGAGAAATCTTGGG 1980
Qy 2357 GGCATATTTTGAATCTCTGCTCCCTCTCTCTCTTCTTGTGTTGTGATCTGATTCAGG 2416
Dh 1981 GGCATATTTTGAATCTCTGCTCCCTCTCTCTCTTCTTGTGTTGTGATCTGATTCAGG 2040
Qy 2417 TGGCCAAAGTATTTGATTTCTTTCAGACAGGCTATATTTGAAGAAACCTTGTATGTG 2476
Dh 2041 TGGCCAAAGTATTTGATTTCTTTCAGACAGGCTATATTTGAAGAAACCTTGTATGTG 2100
Qy 2477 GCTGAAGAAATTTAAGCAGAGCTAGAAAATCATATCAAAATTTTACTGCTCAAGATG 2536
Dh 2101 GCTGAAGAAATTTAAGCAGAGCTAGAAAATCATATCAAAATTTTACTGCTCAAGATG 2160
Qy 2537 GAAACCTCTCTCTTCCAGCTTATGATTTTCTAAGTATGATGATGCGCAACATCTCAACG 2596
Dh 2161 GAAACCTCTCTCTTCCAGCTTATGATTTTCTAAGTATGATGATGCGCAACATCTCAACG 2220
Qy 2597 TGGCAGTATGGAATTTCCCAAGTGTATGAAAGAAACAGATGTGAGAGAAATGATTAAGA 2656
Dh 2221 TGGCAGTATGGAATTTCCCAAGTGTATGAAAGAAACAGATGTGAGAGAAATGATTAAGA 2280
Qy 2657 ACAACATGAGATTAATTCGAGAGTCAGAAAAGAACCATTTGAATCTGAGAAATTAAGCA 2716
Dh 2281 ACAACATGAGATTAATTCGAGAGTCAGAAAAGAACCATTTGAATCTGAGAAATTAAGCA 2340
Qy 2717 TTGGAATAGAAATTAATTTGTTTAAAGAGCAAGAAAATCAATTTAGACGTATGAAGAGTT 2776
Dh 2341 TTGGAATAGAAATTAATTTGTTTAAAGAGCAAGAAAATCAATTTAGACGTATGAAGAGTT 2400

```

OY 2777 AGAGTCGCAAAAGTGAAGTAATTTTATAGGGCTGTGTTCCAAATTTTGGCATG 2836
    |||||||
Db 2401 AGGATGCTCAAAAGTGAAGTAATTTTATAGGGCTGTGTTCCAAATTTTGGCATG 2460
OY 2837 ATGAGCTAATTTATTTTCCCTTAAGAGTAATATTAATCATTTTCAGTTTGCAGCTAGT 2896
    |||||||
Db 2461 ATGAGCTAATTTATTTTCCCTTAAGAGTAATATTAATCATTTTCAGTTTGCAGCTAGT 2520
OY 2897 GCCATCCAAATGAAATTAATTAATAGTCACATTTTATTTAAATTTTCTGTAGTAC 2956
    |||||||
Db 2521 GCCATCCAAATGAAATTAATTAATAGTCACATTTTATTTAAATTTTCTGTAGTAC 2580
OY 2957 ATTTAAACAAAGTAAAGTGAAGGAGGCAAAATATTTTGATATTAATTTTCCACCAAGTAG 3016
    |||||||
Db 2581 ATTTAAACAAAGTAAAGTGAAGGAGGCAAAATATTTTGATATTAATTTTCCACCAAGTAG 2640
OY 3017 TATGCCCAAAATGGAATATAGAAATTAATTAATGAGATNTTTTACATCTTTTGTGA 3076
    |||||||
Db 2641 TATGCCCAAAATGGAATATAGAAATTAATTAATGAGATNTTTTACATCTTTTGTGA 2700
OY 3077 CCAAGTCTTCTAAATGAGTACATATTTTATPACTGATGATTTCTTCTCCGAGTAGC 3136
    |||||||
Db 2701 CCAAGTCTTCTAAATGAGTACATATTTTATPACTGATGATTTCTTCTCCGAGTAGC 2760
OY 3137 CATATTTCAAGTGTTCATTTGCCACATGTGGCTGTGACTACTGTATGAGACAGTTGAGTA 3196
    |||||||
Db 2761 CATATTTCAAGTGTTCATTTGCCACATGTGGCTGTGACTACTGTATGAGACAGTTGAGTA 2820
OY 3197 CTAGACAAAACCTAGCATTAATTAATTAATTTAGTGTCTAGCCATGATTTCTATTTGATTAANT 3256
    |||||||
Db 2821 CTAGACAAAACCTAGCATTAATTAATTAATTTAGTGTCTAGCCATGATTTCTATTTGATTAANT 2880
OY 3257 TAAACTCAATCAACAGTTAACTCCACAGTGCATTCATGAGTGTGACAGTATATTTGTTT 3316
    |||||||
Db 2881 TAAACTCAATCAACAGTTAACTCCACAGTGCATTCATGAGTGTGACAGTATATTTGTTT 2940
OY 3317 TATTGAGTCAATGATATTAATAATGAGCTTTGTCAACCTCAGGGGATATTTAGCAATGT 3376
    |||||||
Db 2941 TATTGAGTCAATGATATTAATAATGAGCTTTGTCAACCTCAGGGGATATTTAGCAATGT 3000
OY 3377 CGGGAGACATTTTGTGATGTCTAGTACAGGAGTATTTAGACTTTTAGTAGAGTAGAGGCCA 3436
    |||||||
Db 3001 CGGGAGACATTTTGTGATGTCTAGTACAGGAGTATTTAGACTTTTAGTAGAGTAGAGGCCA 3060
OY 3437 TGGATCCGCTAAATTAACCTGATTTGAGACAGGAGGCCCAACAAGAAATTTATCCGCGCG 3496
    |||||||
Db 3061 TGGATCCGCTAAATTAACCTGATTTGAGACAGGAGGCCCAACAAGAAATTTATCCGCGCG 3120
OY 3497 AAATGCTAGTGTGCGCAAGGCTGAGTAACCTTGTGTAAAGTAACCTGTGCGACAGTAG 3556
    |||||||
Db 3121 AAATGCTAGTGTGCGCAAGGCTGAGTAACCTTGTGTAAAGTAACCTGTGCGACAGTAG 3180
OY 3557 GTTTCAGAAATTTCTGCTGTCTGCTACAGTATCATGTTTGAAAAATTTTGGCTATTA 3616
    |||||||
Db 3181 GTTTCAGAAATTTCTGCTGTCTGCTACAGTATCATGTTTGAAAAATTTTGGCTATTA 3240
OY 3617 GATATGATTTGATGAGGCTTATTCCTGATTAATTAATCCTGATCAACTGATCTTTCTAT 3676
    |||||||
Db 3241 GATATGATTTGATGAGGCTTATTCCTGATTAATTAATCCTGATCAACTGATCTTTCTAT 3300
OY 3677 ATTTTCAGAAAGTGAATGGAATTAACCTTAGAAGAGAGCTCAGAAATGATATTTATTTTAA 3736
    |||||||
Db 3301 ATTTTCAGAAAGTGAATGGAATTAACCTTAGAAGAGAGCTCAGAAATGATATTTATTTTAA 3360
OY 3737 GTGAGTCTTTAAACCTCTCTTATTTTACAGAGTTATATGGCTAAATTTTCAGATTGAACA 3796
    |||||||
Db 3361 GTGAGTCTTTAAACCTCTCTTATTTTACAGAGTTATATGGCTAAATTTTCAGATTGAACA 3420
OY 3797 GGGATTCAGACATTTGCGCATCTCATAGAAAGAGAGGCTCCCATCTGGAAGGCTGC 3856
    |||||||
Db 3421 GGGATTCAGACATTTGCGCATCTCATAGAAAGAGAGGCTCCCATCTGGAAGGCTGC 3480
OY 3857 TGAATCTACCCCTTGCAAGCTTCAGACAAATCATGTTGATCTTCCCTGAGGCCACAGGCCCTC 3916

```

```

Db 3481 TGAATCTACCCCTTGCAACCTTCAGACAAATAGTTGATCTCCGTGAGGCACACGCCCTC 3540
OY 3917 ATTCTGTGAGGAGGAAAGATTAGCAAAAGGTTAATTTTATTTCAATATCAGTATAGCT 3976
    |||||||
Db 3541 ATTCTGTGAGGAGGAAAGATTAGCAAAAGGTTAATTTTATTTTCAATATCAGTATAGCT 3600
OY 3977 GTTAGACATGATCTGTTTGTAGCAGTGTGTTGTCTCATATTTTCTGCTGTGATTTTGTAG 4036
    |||||||
Db 3601 GTTAGACATGATCTGTTTGTAGCAGTGTGTTGTCTCATATTTTCTGCTGTGATTTTGTAG 3660
OY 4037 ACATTTTGTGAGAAATTTCTATTTGTGCTGTACGTATATTTTCTTTTAAATATCTACT 4096
    |||||||
Db 3661 ACATTTTGTGAGAAATTTCTATTTGTGCTGTACGTATATTTTCTTTTAAATATCTACT 3720
OY 4097 GATATCTGTCTTTAAATTTCTTCAATATGATGTTTGGCTGATTCACATGATTTTATA 4156
    |||||||
Db 3721 GATATCTGTCTTTAAATTTCTTCAATATGATGTTTGGCTGATTCACATGATTTTATA 3780
OY 4157 ACTGAAATTTAAGAAATCTAACAGCTAAACCTCAGTAAGTGCATMTATTTCTTATTAACA 4216
    |||||||
Db 3781 ACTGAAATTTAAGAAATCTAACAGCTAAACCTCAGTAAGTGCATMTATTTCTTATTAACA 3840
OY 4217 TAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTCTTCTGTAGCATGTGATGC 4276
    |||||||
Db 3841 TAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTCTTCTGTAGCATGTGATGC 3900
OY 4277 CTGATTAATCACTATTTCTCATTTGCTTTTATTTCTTAATAGGAAACATGAGTGACATC 4336
    |||||||
Db 3901 CTGATTAATCACTATTTCTCATTTGCTTTTATTTCTTAATAGGAAACATGAGTGACATC 3960
OY 4337 TAAATATAGTGTGTAGTAAATTAACATCATTTAGCCATATTTATAGAAATGCTAATTAAG 4396
    |||||||
Db 3961 TAAATATAGTGTGTAGTAAATTAACATCATTTAGCCATATTTATAGAAATGCTAATTAAG 4020
OY 4397 TACACGACATAGAAACATGAATTTGCTAGTATTTGATGATGATGCTGTGACAAATTTTGACA 4456
    |||||||
Db 4021 TACACGACATAGAAACATGAATTTGCTAGTATTTGATGATGATGCTGTGACAAATTTTGACA 4080
OY 4457 GTCATTAATGTTTGTCTATTAATTTTAAATTAAGTGTGCTGGTTTTCAGAAATACCTTC 4511
    |||||||
Db 4081 GTCATTAATGTTTGTCTATTAATTTTAAATTAAGTGTGCTGGTTTTCAGAAATACCTTC 4135

RESULT 2
US-10-144-771-15372
; Sequence 15372, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: C1001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 15372
; LENGTH: 3756
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-15372

Query Match          20.6%; Score 934.6; DB 8; Length 3756;
Best Local Similarity 63.7%; Pred. No. 2,5e-186;
Matches 1859; Conservative 0; Mismatches 614; Indels 446; Gaps 11;

```

Db	494	ATTAGTTTATTTCGCCAAGTTGGTTCTCTAACTACTCTGTTCTTCTGGAAAGCAATAA		553
Oy	518	TGAGCTGTATTAGACCTCTCGACAGACATATGSAATTTGAGTTGCCAAGTTAATTGTGCAA		577
Db	554	AGAAAGCTGTCAAGCCTCTCCAGGACACTATGGAATTTCAATTGCAAGAGTTACTTGTTGA		613
Oy	578	AGAAAGAAATATCAAGATACGTGTGSAANAAGAAAGAAATTTGATGCAAGCATATTTATTCAA		637
Db	614	AGAAAGAACCCGCAAGGATACGTGTGGGGAAGAGGGTTTGATGCAAGCCATTTATTATCAG		673
Oy	638	GGGCAACATATTTGCTCGAGAAATTTCCCTACCTGACACCTGTGTGATGAAATGCAATGCT		697
Db	674	AGGCACACTACTTCTCAGAGAAATTTCCCAAGGACATCTTGTTGATGTGAAGCCCATCAT		733
Oy	698	CGCCCATGTTCTCTTCTCTCTCTTTTAAAGTGAAGTGAATATATTACCAACCTGGAGA		757
Db	734	TGCTCATGTTCTCTTCTGCTGCTTTTATTAGAAGTAAATATATTACCAACCTGGAGA		793
Oy	758	CCTTTCAGACATATAGAAATAGCTCTGGAAGAGAAAGCAATATATATTCTCATATGTAG		817
Db	794	CCTTTCACGCAATAGAAAACCTCTGGAAGAGAAATATAATGATATTTCTCATAGTGA		853
Oy	818	AGCCATTGGAATTACACAGACACAGACAGATGATGAAAGCCGGTTTGATATGGGACATAC		877
Db	854	AGCCATGCGAACCACACAGATAGACAGATCATGAAAGCTGTTTGTTGATGGGACCTC		913
Oy	878	ATACCAATTTGTTCTTAAACACAGAAATGGCTTTTGGAAAGTATTTGGCTCGAGAGATG		937
Db	914	TTTACAGATTTGCTTAAACACAGAAATTTGCTTTTAAAGAAATATGCGCTCGAGAGAT		973
Oy	938	GGATATATGACATCTCTCATTTTTCATTBTAAACTATGCTTGGACTTACCCACAGATG		997
Db	974	AGAACATATGACATCTCTCACTTTTTCATTGTAAACGGCTTTGGACTTACCCGCAATTG		1033
Oy	998	TAGAAGAACACTAATATGGAACAGCCATATGACTGATCACTGAACATTTACCTGTATTAAGAC		1057
Db	1034	CAGAAGAACACTATATGSAAGAACGACATTTGACACACATCAATATTCACGTATTTGTGAAGAC		1093
Oy	1058	AATGAAGACACCTCTGTTGACTGAAGTGTGTAAGATCTCTCAACAAAGTTTCAACTGTCCA		1117
Db	1094	AATGAAGCACCTCTGTT-----		1153
Oy	1118	TCTCCAACTGGGCTTACCACCTGGTTTTTATTGTTAGCCAAACAGGCTACTTATGAAGCTGA		1177
Db	1112	-----		1213
Oy	1178	TAGAAGAACTGCANAAATGGGTTGCTTGGGCTCTTGGGAAAGCAAGAGTTCCTACTCTT		1237
Db	1112	-----		1273
Oy	1238	GTTAAGGAGACTTTTGGAAATGAACATCTCTCAAGATGCTAATGTGTTCTCAAAAGAC		1297
Db	1112	-----		1333
Oy	1298	AGAAAGGAGATTCACAGTGAATTTTGGTATTTACATGATGTTGAATTTAATATATTCOA		1357
Db	1112	-----		1393
Oy	1358	TGTGGAATAATATATGCACTTGAAGAAATATACAGAAGATGAAAGCAATGACATGGAAGG		1417
Db	1166	CGTAAATAATATATGCACTTGAAGAAATATACAGAAGATGAAAGCGAAGACATGGAAGG		1453
Oy	1418	TTCGCAATATGATGTTTCAGGATGATGAAGGGCAGAACTGTTTTCAGAGATAGGAAGAG		1477
Db	1226	CCCGAGATCTGGCTTTGGAATATGATGAGGGAGGAACTGTTTTCAGAGATAGGAAGAA		1495
Oy	1478	AAATATACCTTTTGGAACTTTACAGTGAACATACAGAAAGAAATTTAATGCAACAGTAT		1537
Db	1286	ACCACTACCTTTTGGAGTTGCTGTGTGACGCTAACAGAGGAGACTTTTAAATACACGCGTAT		1545
Oy	1538	GGCTTTTCAGACGATAGTACTTCTTATGCTGGTTTGGCAAGCAGTATCCATGGCAATTTT		1597

D	b	1346	GATTCTGACAGCAMPATNGTCCTTATAGTGACT-----	1380
O	y	1598	GCATCTCATTTATGATGGCAGTTAAACTGAAAAGCACATCTACTACTGTCTTACTAG	1657
D	b	1381	----- -CCAGTCCACCACATCTTCTCCACAG	1405
O	y	1658	AATAAAGTGGCAGATTGGTGTGATGTACTTAAGCAAAATGTTAGTAATTTCCAT	1712
D	b	1406	AATAAATTGTCCAAATTTGGTGTGATATATGTACGAAGCAAATGTACTGCAATTTCCCGT	1465
O	y	1718	CATAAGATGTACAAGAAAGGCCGAGAACCAGTATCTTATGCTGGAATTTAGAACAA	1777
D	b	1466	GGTAAAGTTGTAATAGAAGGCCGAGACCACAGTGTCTTATGCTGGATATGTGCAACTAA	1525
O	y	1778	AGATCTCTAAATTTATCCAGCTCAACAGGATTTCAITCCAAGTAATATATACATGCAT	1837
D	b	1526	AGATCTCTGAAATTTATCCAGTTGAATAGATCTCTCTCCAGTAGAACATAGCATAT	1585
O	y	1838	CCAGAAGCACAAGAAATTTAAGTGGGGAATATATTAAGAACCTCATCTGTATTCTAG	1897
D	b	1566	CCAGGAAGCACAAGAAATPACCTGCGTGGGGAATATATTAAGAACCTCCAGCTCTGTAG	1645
O	y	1898	TGTGTCAATTTGGGACTATTATTAGTCCAACCATGAAGAAACAGAAAAAGATTTAGTA	1957
D	b	1646	TGTGTCAATTTGGGACTCTTTCACGCCAGCATGGCATCAGCAAAAAGAACTTTTAGGA	1705
O	y	1958	AGCAGAAATACCTTAAGAGATATGTATCACTGGAATTTATCTGAAGAAGATGTTTT	2017
D	b	1706	AGCAGGAAGCAGCTGAGAGGCTGTGATCATACGGAAATTTATCTGAAGATGATGTGTG	1765
O	y	2018	GCTACTGTCAACCAATATGTCTGCAAGTCTTCCAGCCCTGTGCTTCCAGACACACAGA	2077
D	b	1766	GATACTGTCAAAATTAATATCTCAACATCTCCAGCCGTGTACTGCGCCAGGCTTAAGAA	1825
O	y	2078	AGGCAAAATAGAGACATCTCCACTAGCTAGCACATGCAACAAGACATAGTTCAATAT	2137
D	b	1826	AGGTAGATATAGAGTGTCTCCACTAGACACACACACACACACACACACACACACAC	1885
O	y	2138	AACAGATGCACTACTGSAANAATGTTTCCGGAATCACTGTGSAANAATCTTCCAGTTATTT	2197
D	b	1886	AGCAAAATGCACTACTGSAAMCATTTTGG---TGCAGCATTAACCCCTCAGTACAGAAAC	1941
O	y	2198	CAGACTTCAGAAACCATTTATGATTTTGTCACTGATGACACTGTAAATCTCATATATA	2257
D	b	1942	ACGATTCCTGGCACTGGTGAACAGAAACAGACCTGTGATCTTTACCCCTTGGTGAAT	2001
O	y	2258	AAAAGCATATTTGCACTGTGTAAGCAGAAATCTGTGAATCATTTACTCCATGCTGTGT	2317
D	b	2002	CTTAAGAACGATTAGACTACTCAACCATCTTCTTGCA-----	2040
O	y	2318	AAATCTAAAGATATCTCAGTGGGAGAGAAATCTTGCGGCATATTTGATCCTCTGCC	2377
D	b	2041	-----GGAAAGAACCTCCGTTGGGAGAGGGGCATCTTGAAGCATATTTTGGCGCTGCC	2095
O	y	2378	TCCCCCTCTCTCTTGTGTTGGTGAATCTGCATTCAGGTGGCCAAATTTGCAATTTCC	2437
D	b	2096	ACCACCTTCACAGCTGTGTTGGTGAATCTGCATTCGGSTGGCCAAATGTATGATTTCC	2155
O	y	2438	TTTCAGACAGGCTATTAATTAAGAAAACTGTGATTTGTGGCTGAAGAAATTTAGAAGAGG	2497
D	b	2156	TTTCACTCCAGTCTGTGACGGAACAAAGCCTGTGTTGTGGCTGAAGAACTTTACAAGCCGG	2215
O	y	2498	ACTTAAGAAATCATATCCAAATTTTACTCTGTCAAGATGGAAGAACCTCCCTTCCAGGCTTA	2557
D	b	2216	ACTTAAGAAATCCCATCCAGTCTTGTCTGTGCCAGCATGGAACCTCTCTTCTCTGCCCTT	2275
O	y	2558	TGATTTTCTTAAGTATGATTAAGATCCCCCAACATCTCAACCTGGCACATAGAAAGTTCCCA	2617
D	b	2276	TGATTTTCTGAATATATATGATGAGCCTCTCATATCCAGCTGCCAGGAAGAAAGTTCTAGA	2335
O	y	2618	GTTGATTAAGAAAGAACAGATGTGACGAGATGATAAGAACAAATGAAGATTAATGCGC	2677
D	b	2336	GTTGATTAAGAAAGAACAGATGTGACGAGATGCTGTGAGTGTGCTGGAGACATATC	2395

QY	2678	AGTCAGAAAAGAACCCGTTTAAACTCTGTGGAAATTAAGCAATTGGATPAGAAATTTGGTT	2737
Db	2396	AGCCAGAGAGAGACCCGGTGCAAATTCGTGTGAATTAAGCCCTGGAACACTGCCCACTGGCC	2455S
QY	2738	TAAAGAAAGCAGAAAAATCATTTTGAAGCTGTATTAAGAGTTAGGATGCTCAAAAGTGAACTA	2797
Db	2456	CAAGACACACACAGGAGCCCTTTTCCACACAGCAAGAGAGTTATGA-----	2498
QY	2798	ATTTTATAGGCGCTGTGTTTCCAAAATTTTTTGGCATGATAGACTTAATTTATTCCTT	2857
Db	2499	-----GCGGGCGCTGTGGAGGAGCCCTCTGTGGCACAGGAAACCAGATCAAAATCTGTA	2554
QY	2858	AAAGAAATATATTAATCATTTTAACTTTGACAGTACAGTACAGTCCATCCATAGATATATAT	2917
Db	2555	AAGGACACAGGACTTAATTTATGCCA--TCTGTAGACAAAGGCCCATTTGAATGGGCTTGTAT	2612
QY	2918	ATTAAGTCACATATTTTATTTTAAATTTTCTAGTAACATTAACAAAGTAAAGATGAG	2977
Db	2613	ACAAATGAC-----ATGTAATAATTTCTTAGTACTCTTTAAAAACCCGTAACCGAG	2655S
QY	2978	CAGGGCAAAATTAATTTGATTAATTACTTTTCACCAGTACGTAGTATACCAAAATAGCAATA	3037
Db	2666	CAG-----ATTAACCCCAAGACAGCTTAATT	2690
QY	3038	TAGAAATTAATTAAGAAATTTTACATCCCTTTTGTACCAAGCTCTTAATGACASTA	3097
Db	2691	ACTAGCAAAATGTTGACAT-----TTACTTTTATGTGAAGCTTCAAAATACAGTCT	2743
QY	3098	CATATTTTATACGTACTGCTCATTTCTTACGTCGAGTAGCCATATTTCAAGTGTTCATTGC	3157
Db	2744	C--ATTTCACACTTAATGCTTTTCTTCCACTAAGAGAGATCCGTTTTTGAAGGGCTTATTTGC	2801
QY	3158	CACATGTGGCGCTGTGACTACTGTATGTGAGCAGTCTAGTACTAGCAAAAACTAGCATAT	3217
Db	2802	CATGGGGGTACATGACAGTTGTATATGGGCAAGTCTCATATACAGCAACACACAGTGTCT	2861
QY	3218	TAACTTAGTCTAGCCATGATTTCTATTTTGGATTAATAAT	3256
Db	2862	TGACCTGGGCTTAGCCATCTCTGTGACTTTTGANTCAAAATT	2900
RESULT 3			
US-912-293-92043			
Sequence 92043, Application US/09912293			
GENERAL INFORMATION:			
APPLICANT: Rosen, et. al.			
FILE OF INVENTION: Human Genes, Sequences, and Expression Products 100			
FILE REFERENCE: PO-100			
CURRENT APPLICATION NUMBER: US/09/912,293			
CURRENT FILING DATE: 2001-07-26			
PRIOR APPLICATION NUMBER: 08/103,744			
PRIOR FILING DATE: 1993-08-09			
PRIOR APPLICATION NUMBER: 09/249,651			
PRIOR FILING DATE: 1999-02-12			
PRIOR APPLICATION NUMBER: 08/104,507			
PRIOR FILING DATE: 1993-08-09			
PRIOR APPLICATION NUMBER: 08/196,363			
PRIOR FILING DATE: 1994-02-15			
PRIOR APPLICATION NUMBER: 09/859,490			
PRIOR FILING DATE: 2001-05-18			
PRIOR APPLICATION NUMBER: 08/196,362			
PRIOR FILING DATE: 1994-02-15			
PRIOR APPLICATION NUMBER: 08/221,623			
PRIOR FILING DATE: 1994-03-31			
PRIOR APPLICATION NUMBER: 08/220,691			
PRIOR FILING DATE: 1994-03-31			
PRIOR APPLICATION NUMBER: 09/741,830			
PRIOR FILING DATE: 2000-12-22			
PRIOR APPLICATION NUMBER: 09/813,155			
PRIOR FILING DATE: 2001-03-21			
Remaining Prior Application data removed - See File Wrapper or PALM.			
NUMBER OF SEQ ID NOS: 244538			

```

: SEQ ID NO 92043
: LENGTH: 435
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (320)..(320)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (328)..(328)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (349)..(349)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (352)..(352)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (355)..(355)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (360)..(360)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (370)..(370)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (387)..(387)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (390)..(390)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (398)..(398)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (400)..(400)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (402)..(402)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (412)..(412)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (415)..(415)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (433)..(433)
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   OTHER INFORMATION: n is equal to a,t,g, or c
:   US-09-912-293-92043
:
: Query Match 7.8%; Score 354.2; DB 6; Length 435;
: Best local similarity 94.2%; Pred. No. 9.9e-65;
: Matches 373; Conservative 0; Mismatches 22; Indels 1; Gaps 1;
:
: 2716 ATTGGAATAGAGATGTTAGTTTAAAGAGAGCAAGAAATCATTTAGAGTGATTAAGAGT 2775
: ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
: ATTGGGACGACGAGATGTTTAAAGAGAGCAAGAAATCATTTAGAGTGATTAAGAGT 61

```

QY	2776	TAGATGCTCAAAAGTCAACTAATTTTATATGAGCGCTGTGTTTCCAAAATTTTGGCAT	2835
Db	62	TAGATGCTCAAAAGTCAACTAATTTTATATGAGCGCTGTGTTTCCAAAATTTTGGCAT	121
QY	2836	GATGACTTAATTAATTCCTTAAGATPATATTAATCATTCCTCAAGTTGACAGTAG	2895
Db	122	GATGACTTAATTAATTAATTCCTTAAGATPATATTAATCATTCCTCAAGTTGACAGTAG	181
QY	2896	TGCGATCCAAATAGATTAATATATATTAAGTCAATATTTATTAATAATTTTCAGTAAGTA	2955
Db	182	TGCGATCCAAATAGATTAATATATATTAAGTCAATATTTATTAATAATTTTCAGTAAGTA	241
QY	2956	CATTAACAAGAATTAAGTACGACGGCCAAAATATTTTGATATTAATTTCACCCAGTA	3015
Db	242	CATTAACAAGAATTAAGTACGACGGCCAAAATATTTTGATATTAATTTCACCCAGTA	301
QY	3016	GTATACCCAAAATAGCGAAATATAGAAATTA-TTAATGAGATATTTTACATCCTTTTGG	3074
Db	302	GTATACCCAAAATAGCGAAATATAGAAATTAATTAATGCGATTTTTCANCCNTTTTG	361
QY	3075	TACCAAGTCTTCAATAGCATCATATTTTATAC	3110
Db	362	GACCCAGTNTTCTPAAGCGGTCAAAATNTTATAC	397

```

RESULT 4
US-09-513-999C-32911
Sequence 32911, Application US/09513999C
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59.052.REG
CURRENT APPLICATION NUMBER: US/09/513.999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 32911
LENGTH: 355
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15
OTHER INFORMATION: s-g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
OTHER INFORMATION: k-g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 70
OTHER INFORMATION: s-g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: 325
OTHER INFORMATION: r-a or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: 326
OTHER INFORMATION: n-a, g, c or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 333
OTHER INFORMATION: w-a or t
US-09-513-999C-32911
Query Match 7.5%; Score 341; DB 6; Length 355;
Best Local Similarity 98.3%; Pred. No. 5.6e-62;

```

[illegible]

```

RESULT 5
US-09-912-293-91554
; Sequence 91554, Application US/09912293
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 100
; FILE REFERENCE: PO-100
; CURRENT APPLICATION NUMBER: US/09/912,293
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 08/103,744
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 09/249,651
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 08/104,507
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 08/196,363
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 09/859,490
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 08/196,362
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 08/221,623
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 08/220,691
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 09/741,830
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/813,155
; PRIOR FILING DATE: 2001-03-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 244538
; SEQ ID NO 91554
; LENGTH: 313
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (310)..(310)
; OTHER INFORMATION: n is equal to a,t,g, or c
;S-09-912-293-91554

```

Query Match	6.9%	Score 312:	DB 6:	Length 313:
Best Local Similarity	99.7%	Pred. No. 6.7e-56:		
Matches 312:	Conservative	0:	Mismatches 1:	Indels 0:
			Gaps 0:	
1911	GGACATTTTACTCCACCATGAAACGACAAAGACAGATTTTACTGTAGACGAGAAACTTAC	1970		

Db 1 GGACTATTAGTCACACCATGAAACAGCAAGAAAGATTTTAGTGAGGACGGAATCTAC 60
Qy 1971 CTAAGAGATATGTATACGTGGAATTTATCTGAGAGAGATGTTTCTACTCTCAACC 2030
Db 61 CTAAGAGATATGTATACGTGGAATTTATCTGAGAGAGATGTTTCTACTCTCAACC 120
Qy 2031 AATATGCTGCAAGTCTCCAGCCCTGCTGCTGGCAGACACAGAGGCAAAATAGAG 2090
Db 121 AATATGCTGCAAGTCTCCAGCCCTGCTGCTGGCAGACACAGAGGCAAAATAGAG 180
Qy 2091 AGCATCCAGTAGCTAGACACACATGACACAGACATAGTTCAATATATACAGATCAGCTA 2150
Db 181 AGCATCCAGTAGCTAGACACACATGACACAGACATAGTTCAATATATACAGATCAGCTA 240
Qy 2151 CTGGAATGTTCCGGAATATCTGTGGAATCTCCAGTATTTAGACTTACAGAA 2210
Db 241 CTGGAATGTTCCGGAATATCTGTGGAATCTCCAGTATTTAGACTTACAGAA 300
Qy 2211 CCATTATGATTT 2223
Db 301 CCATTATGATTT 313

RESULT 6
US-09-912-293-96911
: Sequence 96911, Application US/09912293
: GENERAL INFORMATION:
: APPLICANT: Rosen, et. al.
: TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 100
: FILE REFERENCE: PO-100
: CURRENT APPLICATION NUMBER: US/09/912, 293
: PRIOR FILING DATE: 2001-07-26
: PRIOR APPLICATION NUMBER: 08/103,744
: PRIOR FILING DATE: 1993-08-09
: PRIOR APPLICATION NUMBER: 09/249,651
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 08/104,507
: PRIOR FILING DATE: 1993-08-09
: PRIOR APPLICATION NUMBER: 08/196,363
: PRIOR FILING DATE: 1994-02-15
: PRIOR APPLICATION NUMBER: 09/859,490
: PRIOR FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 08/196,362
: PRIOR FILING DATE: 1994-02-15
: PRIOR APPLICATION NUMBER: 08/741,830
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 09/813,155
: PRIOR FILING DATE: 2001-03-21
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 244538
: SEQ ID NO 96911
: LENGTH: 496
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (10)..(10)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (160)..(160)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (184)..(184)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature

: LOCATION: (214)..(214)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (232)..(232)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (238)..(238)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (285)..(285)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (287)..(287)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (332)..(333)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (347)..(347)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (353)..(353)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (367)..(367)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (404)..(404)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (407)..(407)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (411)..(411)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (420)..(420)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (426)..(426)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (439)..(439)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (443)..(443)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (445)..(445)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (472)..(472)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (481)..(481)

OTHER INFORMATION: n is equal to a,t,g, or c
US-09-912-293-96911

Query Match	5.9%	Score 265.8;	DB 6;	Length 496;
Best Local Similarity	83.7%	Pred. No. 3.6e-46;		
Matches 406; Conservative	0;	Mismatches 64;	Indels 15;	Gaps 10;

QY	3521	GTACACTGGTGTGTTAAAGTAACTGGGACAGTACAGTTTCCAGATTTCCGTGGTTCGC	3580
QY	3521		3580
Db	13	GTACACTGGTGTGTTAAAGTAACTGGGACAGTACAGTTTCCAGATTTCCGTGGTTCGC	72
QY	3581	TCACGTCATCATGTTTGAAAAAATTTTGGCTATTAAAGATATGATATAGTGGCTCTTATCC	3640
QY	73	TCACGTCATCATGTTTGAAAAAATTTTGGCTATTAAAGATATGATATAGTGGCTCTTATCC	132
QY	3641	TGATATTATACCT -GGATACAACCTTGATCTTTTCT -AATATTTTCAGAAAGTATGGGATTA	3698
Db	133	TGATATTATACCTGGGATATCAACTTGATATTTTGTGAATATTTTTCAGAAAGTATGGGATTA	192
QY	3699	ACCCTA -GAAGAGGACATC -AGATGATATTTTATTTTAA -GTGAGCTTTAAACCTC	3753
Db	193	ACCCTAGGAAGAGGACATCCAGNATATGATATTTTATTTTAAAGTGAAGCTTTAAACCTC	252
QY	3754	CTCTTATTTTCTACAAGTTATATGGCTAAATTTTCAG - -TTGACAGGATTTTCACG -AT	3808
Db	253	CTCTTATTTTCTACAAGTTATATGGGCTAAATATTTAGATTTGGAGACGGGGTTTCACCAATT	312
QY	3809	TCTCCCATCTCTCCAT -GGAAGAAGAGGCTCCCTCATCTGAAGCGTCTCT -GAATCTTAC	3866
Db	313	CTGCCCCATCTCTCCATCTGAGNNAAGAGAGGGCTCTTATCTGNAGGGCTCTGGAATTTTAC	372
QY	3867	CCTTGCAAGCTTCAGACAATTCAGTTGATCTCCCTGAGCACACGGCCCTCATCTCTGAG	3928
Db	373	CCTTGCGAGTTTCAAGCCATCTCAAGTTGGTTTTCGNGGCGACAGGGGCT -ATTCTTTGGG	431
QY	3927	GGAGGGAAAGATTAAGCCAAAGAGTATTAATTTTCATTTCCAATCACATTAGCTGTTTACATGA	3988
Db	432	GGAGGGGAAAGATTAAGCCCAAGGGTATTAATTTTCATTTCCCAANACACTGGGAGTTTACACGA	491
QY	3987	TCTGT 3991	
Db	492	TCTGT 496	

RESULT 7
US-10-349-781-15516
Sequence 15516, Application US/10349781
GENERAL INFORMATION:
APPLICANT: Seilhamer, Jeffrey J.; Delegeane, Angelo M.
APPLICANT: Stunart, Susan G.; Stuve, Laura L.
APPLICANT: Mullaly, Sara J.; Naughton, Rebecca E.
TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE S
FILE REFERENCE: PD-1028-2 CON
CURRENT APPLICATION NUMBER: US/10/349,781
CURRENT FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: 09/540,764
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 08/992,868
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/032,838
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: 08/734,050
PRIOR FILING DATE: 1996-10-18
PRIOR APPLICATION NUMBER: 60/006,111
PRIOR FILING DATE: 1995-10-24
PRIOR APPLICATION NUMBER: 08/733,814
PRIOR FILING DATE: 1996-10-18
PRIOR APPLICATION NUMBER: 60/005,864
PRIOR FILING DATE: 1995-10-26
PRIOR APPLICATION NUMBER: 08/763,920
PRIOR FILING DATE: 1996-12-11
PRIOR APPLICATION NUMBER: 60/008,794
PRIOR FILING DATE: 1995-12-14

```

;
; PRIOR APPLICATION NUMBER: 08/772,783
; PRIOR FILING DATE: 1996-12-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 61458
; SOFTWARE: PERL Program
; SEQ ID NO 15516
; LENGTH: 257
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID NO: hu00067190
US-10-349-781-15516

```

Query Match	Score	257	DB	8	Length	257			
Best Local Similarity	100.0%	Pred. No.	2,2e-44						
Matches	257	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	3845	CTGAAGCGTCTCTGAAATCTACCCCTTGCAAGCTTCAGACAAATCAAGTGTATCCCTGTAG	3904						
Db	1	CTGAAGCGCTCTGAAATCTACCCCTTGCAAGCTTCAGACAAATCAAGTGTATCCCTGTAG	60						
QY	3905	CCACAGCGGCTCATCTGTGAGGAGGGAACATTAGCCCAAGAGTAAATTTCATTCGA	3964						
Db	61	CCACAGGCGTCATCTGTGAGGAGGGAACATTAGCCCAAGAGTAAATTTCATTCGA	120						
QY	3965	AATCACTTAGCTGTAGACTGATCTGTTGTGACAGTGTGTTGTCATATTTTGGCTCTGT	4024						
Db	121	AATCACTTAGCGTGTAGACTGATCTGTTGTGACAGTGTGTTGTCATATTTTGGCTCTGT	180						
QY	4025	GCATTTTTTGAGACATTTGTTGAGATAATTCATATTGGTGCTCTACTGATATTTTCTTTT	4084						
Db	181	GCATTTTTTGAGACATTTGTTGAGATAATTCATATTGGTGCTCTACTGATATTTTCTTTT	240						
QY	4085	TAAATCTCACTGATAT	4101						
Db	241	TAAATCTCACTGATAT	257						

```

RESULT 8
US-10-349-781-36747
; Sequence 36747, Application US/10349781
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.; Delgeane, Angelo M.
; APPLICANT: Stuard, Susan G.; Sluve, Laura L.
; APPLICANT: Mullany, Sara J.; Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE
; FILE REFERENCE: PD-1028-2 CON
; CURRENT APPLICATION NUMBER: US/10/349,781
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 09/540,764
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 08/992,868
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/032,838
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: 08/734,050
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/006,111
; PRIOR FILING DATE: 1995-10-24
; PRIOR APPLICATION NUMBER: 08/733,814
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/005,864
; PRIOR FILING DATE: 1995-10-26
; PRIOR APPLICATION NUMBER: 08/763,920
; PRIOR FILING DATE: 1996-12-11
; PRIOR APPLICATION NUMBER: 60/008,794
; PRIOR FILING DATE: 1995-12-14
; PRIOR APPLICATION NUMBER: 08/772,783
; PRIOR FILING DATE: 1996-12-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 61458
; SOFTWARE: PERL Program

```

SEQ ID NO 36747
LENGTH: 245
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: hu00936261
US-10-349-781-36747

Query Match 5.4%; Score 245; DB 8; Length 245;
Best Local Similarity 100.0%; Pred. No. 7.3e-42;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3845 CTGAAGCTCTCTGAAATCTACCTTGCAGCTTCAGACAAATCAGTTGATCTCCCTGAG 3904
DB 1 CTGAAGCTCTCTGAAATCTACCTTGCAGCTTCAGACAAATCAGTTGATCTCCCTGAG 60
QY 3905 CCACAGGCGCTCATTCTGTGAGGAGGAGGAAGATTAGCCAAAGATTATTTTCATTCCA 3964
DB 61 CCACAGGCGCTCATTCTGTGAGGAGGAGGAAGATTAGCCAAAGATTATTTTCATTCCA 120
QY 3965 AATCAGCTAGCTGTAGACTGATCTGTTGTAGCAGTTGTTGTCTCATTTTGTCTGT 4024
DB 121 AATCAGCTAGCTGTAGACTGATCTGTTGTAGCAGTTGTTGTCTCATTTTGTCTGT 180
QY 4025 GCATTTTGTGAGACATTTGTGAGAAATATTCATTTGTGCTCTACTGATTTTCTTTT 4084
DB 181 GCATTTTGTGAGACATTTGTGAGAAATATTCATTTGTGCTCTACTGATTTTCTTTT 240
QY 4085 TAAAT 4089
DB 241 TAAAT 245

RESULT 9
US-10-349-781-24291
Sequence 24291, Application US/10349781
GENERAL INFORMATION:
APPLICANT: Seilhamer, Jeffrey J.; Deleage, Angelo M.
APPLICANT: Stuart, Susan G.; Stuve, Laura L.
APPLICANT: Mulahy, Sara J.; Naughton, Rebecca E.
TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE
FILE REFERENCE: PD-1028-2 CON
CURRENT APPLICATION NUMBER: US/10/349,781
PRIOR APPLICATION NUMBER: 2003-01-21
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 08/992,868
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/032,838
PRIOR FILING DATE: 1996-12-13
PRIOR APPLICATION NUMBER: 08/734,050
PRIOR FILING DATE: 1996-10-18
PRIOR APPLICATION NUMBER: 60/006,111
PRIOR FILING DATE: 1995-10-24
PRIOR APPLICATION NUMBER: 08/733,814
PRIOR FILING DATE: 1996-10-18
PRIOR APPLICATION NUMBER: 60/005,864
PRIOR FILING DATE: 1995-10-26
PRIOR APPLICATION NUMBER: 08/763,920
PRIOR FILING DATE: 1996-12-11
PRIOR APPLICATION NUMBER: 60/008,794
PRIOR FILING DATE: 1995-12-14
PRIOR APPLICATION NUMBER: 08/772,783
PRIOR FILING DATE: 1996-12-23
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 61458
SOFTWARE: PERL Program
SEQ ID NO 24291
LENGTH: 228
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: hu00693953
US-10-349-781-24291

Query Match 5.0%; Score 228; DB 8; Length 228;
Best Local Similarity 100.0%; Pred. No. 2.6e-38;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3667 CTTTCTATATTTTCAGAAAGTATGGATACCTTGAAGAGACCTCAGATATAT 3726
DB 1 CTTTCTATATTTTCAGAAAGTATGGATACCTTGAAGAGACCTCAGATATAT 60
QY 3727 TATATTTTAAAGTACTTAAACCTCCTTATTTCTCAAGTATATAGCTAAATTC 3786
DB 61 TATATTTTAAAGTACTTAAACCTCCTTATTTCTCAAGTATATAGCTAAATTC 120
QY 3787 AGATTGAACAGGAGATTACGATTTGCCATCTCTCATGAGAAAGAGAGGCTCCATCT 3846
DB 121 AGATTGAACAGGAGATTACGATTTGCCATCTCTCATGAGAAAGAGAGGCTCCATCT 180
QY 3847 GAAGGCTCTGAAATCTACCTTGCAGCTTCAGACAAATCAGTTGA 3894
DB 181 GAAGGCTCTGAAATCTACCTTGCAGCTTCAGACAAATCAGTTGA 228

RESULT 10
US-09-912-293-18131
Sequence 18131, Application US/09912293
GENERAL INFORMATION:
APPLICANT: Rosen, et. al.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 100
FILE REFERENCE: PO-100
CURRENT APPLICATION NUMBER: US/09/912,293
PRIOR APPLICATION NUMBER: 2001-07-26
PRIOR FILING DATE: 1993-08-09
PRIOR APPLICATION NUMBER: 09/249,651
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 08/104,507
PRIOR FILING DATE: 1993-08-09
PRIOR APPLICATION NUMBER: 08/196,363
PRIOR FILING DATE: 1994-02-15
PRIOR APPLICATION NUMBER: 09/859,490
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 08/196,362
PRIOR FILING DATE: 1994-02-15
PRIOR APPLICATION NUMBER: 08/221,623
PRIOR FILING DATE: 1994-03-31
PRIOR APPLICATION NUMBER: 08/220,691
PRIOR FILING DATE: 1994-03-31
PRIOR APPLICATION NUMBER: 09/741,830
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 09/813,155
PRIOR FILING DATE: 2001-03-21
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 244538
SEQ ID NO 18131
LENGTH: 375
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (172)..(172)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (201)..(201)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (223)..(223)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:

```

NAME/KEY: misc_feature
LOCATION: (236)..(236)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (239)..(239)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (242)..(242)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (267)..(267)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (275)..(275)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (304)..(304)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (306)..(306)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (327)..(327)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (339)..(339)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (370)..(370)
OTHER INFORMATION: n is equal to a,t,g, or c
US-09-912-293-18131

```

```

Query Match
Best Local Similarity 92.4%; Score 188.6; DB 6; Length 375;
Matches 218; Conservative 0; Mismatches 14; Indels 4; Gaps 2;

```

```

QY 777 GCTCTGAAGAAAGCAATATATATCTCATATATGAAGCCATTGGAATCCAGAG 836
      |||
DB 8 GCCTGAAAGAAAGCAATATATATCTCATATATGAAGCCATTGGAATCCAGAG 67
QY 837 CACAGAGCATGATGAGACCGCTTTGATGATGAGCATACATACCAATTTGCTTAAC 896
      |||
DB 68 CACAGAGCATGATGAGACCGCTTTGATGATGAGCATACATACCAATTTGCTTAAC 127
QY 897 ACAGAAATTCCTTTGGAAGTATGCTGAGAGATGAG--AATATGACATCTC 954
      |||
DB 128 ACAGAAATTCCTTTGGAAGTATGCTGAGAGATGAGGATGAGGATGAGGATGAG 187
QY 955 ACTTTTTCATGTAACACTAGTCTTGG--ACTTGACCAAGCATATGAGAAGAC 1008
      |||
DB 188 ACTTTTTCATGTAACACTAGTCTTGG--ACTTGACCAAGCATATGAGAAGAC 243

```

```

RESULT 11
US-09-539-806B-30366
; Sequence 30366, Application US/09539806B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Deleane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF EXOCRINE GLAND TISSUE

```

```

FILE REFERENCE: PD-1027 CIP
CURRENT APPLICATION NUMBER: US/09/539,806B
CURRENT FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 08/706,766
PRIOR FILING DATE: September 27, 1996
PRIOR APPLICATION NUMBER: 60/004,676
PRIOR FILING DATE: October 2, 1995
PRIOR APPLICATION NUMBER: 08/749,515
PRIOR FILING DATE: November 15, 1996
PRIOR APPLICATION NUMBER: 60/006,810
PRIOR FILING DATE: November 15, 1995
PRIOR APPLICATION NUMBER: 08/822,285
PRIOR FILING DATE: March 20, 1997
PRIOR APPLICATION NUMBER: 60/013,696
PRIOR FILING DATE: March 20, 1996
PRIOR APPLICATION NUMBER: 08/951,197
PRIOR FILING DATE: October 1, 1997
PRIOR APPLICATION NUMBER: 60/027,249
PRIOR FILING DATE: October 1, 1996
PRIOR APPLICATION NUMBER: 08/826,438
PRIOR FILING DATE: March 20, 1997
PRIOR APPLICATION NUMBER: 60/016,145
PRIOR FILING DATE: April 18, 1996
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 48372
SOFTWARE: PERL Program
SEQ ID NO 30366
LENGTH: 181
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: hu00862945
US-09-539-806B-30366

```

```

Query Match
Best Local Similarity 100.0%; Score 181; DB 5; Length 181;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2756 ATTTAGACGTGATTAAGAGTATGATGCTCAAAAGTGAACATATTTATAGGCGTGTGT 2815
      |||
DB 1 ATTTAGACGTGATTAAGAGTATGATGCTCAAAAGTGAACATATTTATAGGCGTGTGT 60
QY 2816 TTCCAAATTTTGGCATGATGACTTAATTTATTTCTTAAGCAATATTAATC 2875
      |||
DB 61 TTCCAAATTTTGGCATGATGACTTAATTTATTTCTTAAGCAATATTAATC 120
QY 2876 ATTTCAAGTTGAGACTGATGCTGCATCCAAATGAATTAATATAGTACATATTTAT 2935
      |||
DB 121 ATTTCAAGTTGAGACTGATGCTGCATCCAAATGAATTAATATAGTACATATTTAT 180
QY 2936 T 2936
DB 181 T 181

```

```

RESULT 12
US-10-349-781-1729
; Sequence 1729, Application US/10349781
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.; Deleane, Angelo M.
; APPLICANT: Stuart, Susan G.; Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.; Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE
; FILE REFERENCE: PD-1028-2 CON
; CURRENT APPLICATION NUMBER: US/10/349,781
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 09/540,764
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 08/992,868
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/032,838
; PRIOR FILING DATE: 1996-12-13

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2003, 04:56:46 ; Search time 6593 Seconds
(without alignments)
11117.970 Million cell updates/sec

Title: US-09-847-046-1
Perfect score: 4526
Sequence: 1 gccgcagtcgctgtgcgctgc.....ccttcaaaaaaaaaaaaaa 4526

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	776.2	17.1	784	9 AU124110 AU124110
2	714.4	15.8	847	12 BG546766 602574136
3	710.2	15.7	753	14 BM728814 UT-E-E01-
4	702.8	15.5	743	9 AU119721 AU119721
5	688.8	15.2	1266	11 AK017582 Mus muscu
6	639.2	14.1	1453	13 BM552780 AGENCOURT

7	602	13.3	605	10	AV646118	
c	583.6	12.9	593	13	BT172516	
8	581.2	12.8	612	10	AW961708	
10	568.2	12.6	686	13	BG928131	
c	567	12.5	585	14	BM972493	
c	563.6	12.5	564	9	AL040840	
13	562	12.4	900	12	BE895218	
14	556.4	12.3	695	12	BF790996	
c	552.6	12.2	578	10	AW665394	
15	549.8	12.1	578	10	AV662311	
c	549.8	12.1	512	10	AW131594	
c	497	11.0	497	9	AA934775	
c	18	489.6	10.8	497	12	BG863191
19	485.8	10.7	751	12	BT172755	
20	483.4	10.7	488	13	BT172755	
21	481.6	10.6	627	9	AA113856	
22	475.4	10.5	477	9	AA278257	
23	475	10.5	571	12	BG609542	
c	24	469.6	10.4	470	9	AI333000
25	463.8	10.2	728	13	BI689134	
c	26	461	10.2	470	9	AI832055
c	27	441.6	9.8	442	9	AA278675
c	28	439.4	9.7	475	10	AW131649
c	29	427	9.4	439	9	AI306463
c	30	425.6	9.4	426	9	AI453818
c	31	423	9.3	425	9	AI014524
c	32	421.2	9.3	604	9	AA113931
c	33	412.6	9.1	539	14	H06836
c	34	410.6	9.1	415	10	AW150811
c	35	408	9.0	887	14	BQ919809
c	36	407.6	9.0	416	9	AI140821
c	37	407.2	9.0	807	13	BI655128
c	38	398.6	8.8	1077	13	BI689253
c	39	383	8.5	383	9	AA773166
c	40	376.8	8.3	410	9	AI333002
c	41	368.4	8.1	385	9	AA737677
c	42	355.6	7.9	1177	14	BQ215516
c	43	329.6	7.3	588	14	BM941478
c	44	329.6	7.3	604	12	BF458484
45	328	7.2	603	12	BF458491	

ALIGNMENTS

RESULT 1
LOCUS AU124110 784 bp mRNA linear EST 01-AUG-2002
DEFINITION AU124110 NT2RM2 Homo sapiens CDNA clone NT2RM2001681 5', mRNA
ACCESSION AU124110
VERSION AU124110.1 GI:10948826
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 784)
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
TITLE HRI human CDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomicehri.co.jp
HRI human CDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; CDNA library construction; Department of

Virology, Institute of Medical Science, University of Tokyo, and
 Hellix Research Institute.
 Location/Qualifiers
 source
 1.784
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RM2001681"
 /clone_lib="NT2RM2"
 /cell_type="teratocarcinoma"
 /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"

BASE COUNT 262 a 149 c 148 g 222 t 3 others
 ORIGIN

Query Match 17.1%; Score 776.2; DB 9; Length 784;
 Best Local Similarity 99.2%; Pred. No. 3.9e-137;
 Matches 778; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1686 TGTACTAGCAAAATGTACTGTAATTTCTATCATTAAGATGTACAGAAAGCGAGAAC 1745
 |||||||
 Db 1 TGTACTAGCAAAATGTACTGTAATTTCTATCATTAAGATGTACAGAAAGCGAGAAC 60

QY 1746 CCGATATCTTAACTGGTAATGTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 1805
 |||||||
 Db 61 CCGATATCTTAACTGGTAATGTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 120

QY 1806 AGGATTTCAATCCAGTGAATATACATGCATCCAGAAAGCAAGAAATTTAAAGTGG 1865
 |||||||
 Db 121 AGGATTTCAATCCAGTGAATATACATGCATCCAGAAAGCAAGAAATTTAAAGTGG 180

QY 1866 GAATTAATTAAGACCTCATCTTGTATCTAGTGTCTAGTATTTGGACTATTTAGTCCA 1925
 |||||||
 Db 181 GAATTAATTAAGACCTCATCTTGTATCTAGTGTCTAGTATTTGGACTATTTAGTCCA 240

QY 1926 ACCATGAAGCAAGCAAGATTTAGTGAAGCAAGAACTCAATGAAGATATGT 1985
 |||||||
 Db 241 ACCATGAAGCAAGCAAGATTTAGTGAAGCAAGAACTCAATGAAGATATGT 300

QY 1986 ATCACTGAATTTATCTGAGAGAGATGTTTCTACTGTCAACCAATATGCTGCAAGT 2045
 |||||||
 Db 301 ATCACTGAATTTATCTGAGAGAGATGTTTCTACTGTCAACCAATATGCTGCAAGT 360

QY 2046 CTTCACGCCCTGCTGCTCCAGACACAGAGCAAGAAATAGAGACATCCCATAGCT 2105
 |||||||
 Db 361 CTTCACGCCCTGCTGCTCCAGACACAGAGCAAGAAATAGAGACATCCCATAGCT 420

QY 2106 AGACACATGCAACAGATAGTCAATTAATTAACAGATGCACTAGTGAATGTTCCG 2165
 |||||||
 Db 421 AGACACATGCAACAGATAGTCAATTAATTAACAGATGCACTAGTGAATGTTCCG 480

QY 2166 GAATTCATGTTGGAATATCTCCAGTATTTTCAGACTTCAGAAACCATTTATGATTTTG 2225
 |||||||
 Db 481 GAATTCATGTTGGAATATCTCCAGTATTTTCAGACTTCAGAAACCATTTATGATTTTG 540

QY 2226 TTCACTGATGCACTGTAAATCTCAATATTAAGAAACCAATTTGACACTGTAAACGAG 2285
 |||||||
 Db 541 TTCACTGATGCACTGTAAATCTCAATATTAAGAAACCAATTTGACACTGTAAACGAG 600

QY 2286 AAATCTTGATTTCAATTTCTCATGCTGTAAATCTCAAGAAATCTCCAGTGGGAGAG 2345
 |||||||
 Db 601 AAATCTTGATTTCAATTTCTCATGCTGTAAATCTCAAGAAATCTCCAGTGGGAGAG 660

QY 2346 GGAATTCCTGGGCAATTTTGTATCTGCTCCCTCCCTCTCTGTTGTTGGTGAAT 2405
 |||||||
 Db 661 GGAATTCCTGGGCAATTTTGTATCTGCTCCCTCCCTCTCTGTTGTTGGTGAAT 720

QY 2406 CTGCATTCAGGTGGCCAGATTTTGCATTTCTTCAGACAGGCTATTAATGAAGAAAC 2465
 |||||||
 Db 721 CTGCATTCAGGTGGCCAGATTTTGCATTTCTTCAGACAGGCTATTAATGAAGAAAC 780

QY 2466 CTTG 2469
 |||||

Db 781 CTTG 784

RESULT 2
 BG546766 847 bp mRNA linear EST 04-APR-2001
 LOCUS 602574136F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702136 5',
 DEFINITION mRNA sequence.
 ACCESSION BG546766
 VERSION BG546766.1 GI:13545431
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 847)
 AUTHORS NIH-MGC <http://img.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM1538 row: h column: 09
 High quality sequence stop: 809.

FEATURES
 source
 Location/Qualifiers
 1..847
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4702136"
 /clone_lib="NIH_MGC_77"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgccgtcgcc); Site_2: SfiI (ggccatcagcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATTAATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAAGAGCCGAGGCGCCAGCAGT-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 251 a 156 c 158 g 282 t
 ORIGIN

Query Match 15.8%; Score 714.4; DB 12; Length 847;
 Best Local Similarity 94.1%; Pred. No. 2e-125;
 Matches 798; Conservative 0; Mismatches 41; Indels 9; Gaps 5;

QY 3027 ATAGCAAAATATGAATTAATGATATTTTACATCCTTTTGTGACAGCTTC 3086
 |||||||
 Db 1 ATAGCAAAATATGAATTAATGATATTTTACATCCTTTTGTGACAGCTTC 60

QY 3087 TAAATGACATATTTATTTACTCTGATTTCTTACTTCGATGACATATTTCA 3146
 |||||||
 Db 61 TAAATGACATATTTATTTATTTACTCTGATTTCTTACTTCGATGACATATTTCA 120

QY 3147 GTGTCATTTGCCACATGTGGCCGTGACTACTGTATTTGACAGTTCAGTACTAGACAAA 3206
 |||||||
 Db 121 GTGTCATTTGCCACATGTGGCCGTGACTACTGTATTTGACAGTTCAGTACTAGACAAA 180

QY 3207 ACTAGCATTAATTAATTTAGTTTACCATGATTTTCTATTTGATTAATTAACCTTA 3266
 |||||||
 Db 181 ACTAGCATTAATTAATTTAGTTTACCATGATTTTCTATTTGGATTAATTAACCTTA 240

QY 3267 TCACAGTTAACTCCACAGGCAATTCATGACGCGAGAGTATATTTGTTTATTTGGAGTC 3326
 |||||||
 Db 241 TCACAGTTAACTCCACAGGCAATTCATGACGCGAGAGTATATTTGTTTATTTGGAGTC 300

QY 3327 ATGATATTAATCAGCGTTTGCAACCTCAGGGGATATTAGCAATTGTGGGAGACAT 3386
 |||||||
 Db 301 ATGATATTAATCAGCGTTTGCAACCTCAGGGGATATTAGCAATTGTGGGAGACAT 360
 |||||||
 QY 3387 TTTTGATGTCATGACTAGGCGAGTTATAGCATTTAGTAGAGCCATGATCCTGC 3446
 |||||||
 Db 361 TTTTGATGTCATGACTAGGCGAGTTATAGCATTTAGTAGAGCCATGATCCTGC 420
 |||||||
 QY 3447 TAAATTAACCTGATGAGCAGCGCCCAACAAGAATTAATCTCCGCCGAATGTAAGT 3506
 |||||||
 Db 421 TAAATTAACCTGATGAGCAGCGCCCAACAAGAATTAATCTCCGCCGAATGTAAGT 480
 |||||||
 QY 3507 CGTCCGAAGCTGATGATACCTGTGTTAAAGTAACTGTGCGAGACATGATGATCCAGAA 3566
 |||||||
 Db 481 CGTCCGAAGCTGATGATACCTGTGTTAAAGTAACTGTGCGAGACATGATGATCCAGAA 540
 |||||||
 QY 3567 TTTCCGTGCTGCTCAGCATATCATGTTTG--AAAAAATTTTGGCTATTAAAGATATGT- 3623
 |||||||
 Db 541 TTTCCGTGCTGCTCAGCATATCATGTTTG--AAAAAATTTTGGCTATTAAAGATATGT 600
 |||||||
 QY 3624 ATTAGATGCTTTATCCGATTTATCTGATACACTTATCTTTCTTA--TATTTT 3681
 |||||||
 Db 601 ATTAGATGCTTTATCCGATTTATCTGATACACTTATCTTTCTTA--TATTTT 660
 |||||||
 QY 3682 CAGAAAGTATGATTAACCTAGAGAGACCTAGAAATGATTTATTTAATTTAAGTAG 3741
 |||||||
 Db 661 ACAGAAAGTATGATTAACCTAGAGAGACCTAGAAATGATTTATTTAATTTAAGTAG 720
 |||||||
 QY 3742 TCT--TAAACCTCTCTTATTTCTACAACTATATGCTTAAATTTGATTTGAACAG 3798
 |||||||
 Db 721 AGTCTGTAACAACTCTCTTATTTCTACAACTATATGCTTAAATTTGATTTGAACAG 780
 |||||||
 QY 3799 GATTCACATCTCCCATCTCTCATGAAAGAGAGCTCCCTATCTGAAGCTCTGTG 3858
 |||||||
 Db 781 ATTCGGCATCTCTCATCTCTCATGAAAGAGAGGGTCCCTCATCTGAAG- GTTCTG 839
 |||||||
 QY 3859 AATCTAC 3866
 |||||||
 Db 840 AATCTAC 847

RESULT 3
 BM728814
 LOCUS BM728814 753 bp mRNA linear EST 01-MAR-2002
 DEFINITION UT-E-BO1-ai-v-d-13-0-UT.1 UT-E-BO1 Homo sapiens cDNA clone
 UT-E-BO1-ai-v-d-13-0-UT 5', mRNA sequence.
 ACCESSION BM728814
 VERSION BM728814.1 GI:19050147
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 753)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 9704447
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mssoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA
 sequence: 1-116 >MER33#DNA/MERL_type (matched compliment) 267-442.
 >MER20#DNA/MERL_type
 Seq primer: M13 reverse.

FEATURES
 source
 location/Qualifiers
 1..753

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UT-E-BO1-ai-v-d-13-0-UT"
 /clone_110="UT-E-BO1"
 /tissue_type="fetal eye"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UT-E-BO1 is a normalized cDNA library containing the
 following tissue(s): fetal eye. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT73-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 CCGCTATACC. This library was created for the program, Gene
 Discovery in the Visual System, supported by National Eye
 Institute (NEI)."
 BASE COUNT 221 a 137 c 143 g 250 t 2 others
 ORIGIN

Query Match 15 74: Score 710.2; DB 14; Length 753;
 Best Local Similarity 98.6%; Pred. No. 1.3e-124;
 Matches 748; Conservative 0; Mismatches 5; Indels 6; Gaps 3;

QY 3079 AAGTCTTAATGAGACATATTTTATCTGATGATTTCTTACTTCCGATGACCA 3138
 |||||||
 Db 1 AAGTCTTAATGAGACATATTTTATCTGATGATTTCTTACTTCCGATGACCA 60
 |||||||
 QY 3139 TATTTCAAGTGTTCATGTCACATGTGGCCTGTGACTGATTTGACAGTTCAGTACT 3198
 |||||||
 Db 61 TATTTCAAGTGTTCCTGTCACATGTGGCCTGTGACTGATTTGACAGTTCAGTACT 120
 |||||||
 QY 3139 AGACAAACATGACATATTTTACTTACTTACTGATTTCTTACTTCCGATGACCA 3258
 |||||||
 Db 121 AGACAAACATGACATATTTTACTTACTTACTGATTTCTTACTTCCGATGACCA 180
 |||||||
 QY 3259 AACTCTAATCAGATTAATCACTGATGATGATGATGATGATGATGATGATGATGAT 3318
 |||||||
 Db 181 AACTCTAATCAGATTAATCACTGATGATGATGATGATGATGATGATGATGATGAT 240
 |||||||
 QY 3319 TTGGAGTCATGATTTAAATCAGCGTTTGTCAACCTCAGGGGATTTTAAATTTGTCG 3378
 |||||||
 Db 241 TTGGAGTCATGATTTAAATCAGCGTTTGTCAACCTCAGGGGATTTTAAATTTGTCG 300
 |||||||
 QY 3379 GGAGCATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3438
 |||||||
 Db 301 GGAGCATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
 |||||||
 QY 3439 GATCCTGTAATTAATCACTGATGATGATGATGATGATGATGATGATGATGATGATG 3498
 |||||||
 Db 361 GATCCTGTAATTAATCACTGATGATGATGATGATGATGATGATGATGATGATGATG 420
 |||||||
 QY 3499 ATGTAAGTCGCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 3558
 |||||||
 Db 421 ATGTAAGTCGCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 476
 |||||||
 QY 3559 TTCCAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3618
 |||||||
 Db 477 TTCCAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 536
 |||||||
 QY 3619 TATGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3678

|||||
 Db 537 TANGTATTAGATGCTTTATCCGATTAATACCTGATACACCTGATCTTTCTATAT 596
 Oy 3679 TTTCAGAAAGTATGGATTAACCTGAGAGAGAGATCAGATGATATTTATTTAACT 3738
 Db 597 TTTCAGAAAGTATGGATTAACCTGAGAGAGAGATCAGATGATATTTATTTAACT 654
 Oy 3739 GAGCTTTAAACCTCTCTTATTTTACAAATTAATGCTAAATTTACATTTGAACAG 3798
 Db 655 GAGCTTTAAACCTCTCTTATTTTACAAATTAATGCTAAATTTACATTTGAACAG 714
 Oy 3799 GATTGACATTTGCGCATCTCTCATGGAAGAGAGCT 3837
 Db 715 GATTGACATTTGCGCATCTCTCATGGAAGAGAGCT 753

RESULT 4
 AUI19721 AUI19721 743 bp mRNA linear EST 01-AUG-2002
 LOCUS AUI19721 HEMBA1 Homo sapiens CDNA clone HEMBA1006479 5', mRNA
 DEFINITION sequence.
 ACCESSION AUI19721
 VERSION AUI19721.1 GI:10934956
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 Ota.T., Nishikawa.T., Suzuki.Y., Ishi.S., Saito.K., Kawai.Y.,
 Yamamoto.O., Wakamatsu.A., Nakamura.Y., Nagai.T., Sugano.S. and
 Isogai.T.

REFERENCE
 AUTHORS HRI human CDNA project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human CDNA project, 5'- & 3'-end one pass sequencing: Helix
 Research Institute; CDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 Location/Qualifiers
 FEATURES
 source 1..743
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1006479"
 /clone_lib="HEMBA1"
 /tissue_type="whole embryo, mainly head"
 /tissue="embryo, 10 weeks"
 /dev_stage="embryo, 10 weeks"
 /note="Vector: pME18SFL3"
 BASE COUNT 197 a 141 c 147 g 254 t 4 others
 ORIGIN

Query Match 15.5%; Score 702.8; DB 9; Length 743;
 Best Local Similarity 99.4%; Pred. No. 3.3e-123;
 Matches 715; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Oy 3541 ACCTGGCAGACTAGGTTTCCAGAAATTCCTGGTTCTGCTACGATATGTTGAAAA 3600
 Db 181 ACCTGGCAGACTAGGTTTCCAGAAATTCCTGGTTCTGCTACGATATGTTGAAAA 240
 Oy 3601 AATTTGGCATTAAGATATGATTA-GATGCTTATTCCTGATATTAACCTGGATTA 3659
 Db 241 AATTTGGCATTAAGATATGATTAAGATGATGCTTATTCCTGATATTAACCTGGATTA 300
 Oy 3660 ACTTGATCTTTCTTATTAATTTTACAGAAAGTATGAGATTAACCTAGAAAGAGACTAGAA 3719
 Db 301 ACTTGATCTTTCTTATTAATTTTACAGAAAGTATGAGATTAACCTAGAAAGAGACTAGAA 360
 Oy 3720 TGATATTTATTTTAAAGTACTTAAACCTCTCTTATTTTCAAGATATATGCT 3779
 Db 361 TGATATTTATTTTAAAGTACTTAAACCTCTCTTATTTTCAAGATATATGCT 420
 Oy 3780 AATTTACATTTGAACAGGATTCAGCATTCCTCCATTCCTCATGGAAGAGAGCTCC 3839
 Db 421 AATTTACATTTGAACAGGATTCAGCATTCCTCCATTCCTCATGGAAGAGAGCTCC 480
 Oy 3840 CTCATCTGAAGCTCTCTGAAATCTACCTTGCACCTTCAGACAATGATGATCTCC 3899
 Db 481 CTCATCTGAAGCTCTCTGAAATCTACCTTGCACCTTCAGACAATGATGATCTCC 540
 Oy 3900 CTGAGCCACAGGCTCATTTCTGTGAGGAGGAGGAAAGATTAGCCAAAGATTAATTTCA 3959
 Db 541 CTGAGCCACAGGCTCATTTCTGTGAGGAGGAGGAAAGATTAGCCAAAGATTAATTTCA 600
 Oy 3960 TTCCAAATCAGTACGCTGTAGACATCTGTTGAGAGAGTTGCTCATTTTTCG 4019
 Db 601 TTCCAAATCAGTACGCTGTAGACATCTGTTGAGAGAGTTGCTCATTTTTCG 660
 Oy 4020 TCTGTGATTTTGTGAGACATTTGTGAGAAATTTCTATTTGGTCTCACTGATTTT 4078
 Db 661 TCTGTGATTTTGTGAGACATTTGTGAGAAATTTCTATTTGGTCTCACTGATTTT 719

RESULT 5
 AK017582 1266 bp mRNA linear HTC 19-JAN-2002
 LOCUS AK017582
 DEFINITION Mus musculus 8 days embryo whole body CDNA, RIKEN full-length
 enriched library, clone:5730420B22:homolog to KIAA1344 PROTEIN
 (FRAGMENT), full insert sequence.
 ACCESSION AK017582
 VERSION AK017582.1 GI:12856892
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) 8 days embryo CDNA to mRNA,
 clone_lib:RIKEN full-length enriched mouse CDNA library
 clone:5730420B22.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCES
 AUTHORS Carninci.P. and Hayashizaki.Y.
 TITLE High-efficiency full-length CDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCES
 AUTHORS Carninci.P., Shibata.Y., Hayatsu.N., Sugahara.Y., Shibata.K.,
 Itoh.K., Konno.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length CDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE
 MEDLINE 11042159
 PUBMED 11042159
 AUTHORS Shibata.K., Itoh.M., Aizawa.K., Nagaoka.S., Sasaki.N., Carninci.P.,
 Konno.H., Akiyama.J., Nishi.K., Kitsuana.T., Tashiro.H., Itoh.M.,
 Sumi.N., Ishi.Y., Nakamura.S., Hazama.M., Nishino.T., Harada.A.,
 Yamamoto.R., Matsumoto.H., Sakaguchi.S., Ikegami.T., Kashiwagi.K.,
 Fujiwara.S., Inoue.K., Togawa.Y., Izawa.M., Ohara.E., Watanabe.M.,
 Yoneda.Y., Ishikawa.T., Ozawa.K., Tanaka.T., Matsura.S., Kawai.J.,

TITLE
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 Riken integrated sequence analysis (RISA) system--334-Form
 sequencing pipeline with 384 multiplexillary sequencer
 JOURNAL
 Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE
 20530913
 PUBMED
 11076861
 AUTHORS
 4
 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii,

AUTHORS
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, T., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kodota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischnann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, T., Peesle, G., Quackenbush, J., Schriml, L.M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barish, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M., J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinich, S., Hill, D., Hofmann, M., Hume, D.A., Kamliya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, T., Mombeers, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakemoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.C., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Koutsouki, S. and Hayashizaki, Y.

TITLE
Functional annotation of a full-length mouse cDNA collection

JOURNAL
Nature 409 (6821), 685-690 (2001)

MEDLINE
21085660

JOURNAL
MEDLINE
COMMENT

with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
Contact: Zengqiang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzq@chc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
SOURCE

Location/Qualifiers
1. 605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCAJF06"
/clone_lib="GLC"
/issue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 173 a 131 c 156 g 142 t 3 others
ORIGIN

Query Match 13.3%; Score 602; DB 10; Length 605;
Best Local Similarity 99.7%; Pred. No. 4.5e-104;

Matches 602; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GCCGAGTCGTGGCGGCTGACAGGCTGGAGGAGAGAGTCTACGCTTGGCAGTTGGCG 60
1 GCCGAGTCGTGGCGGCTGACAGGCTGGAGGAGAGAGTCTACGCTTGGCAGTTGGCG 60
61 AAGTGTTCACAGGCTACCCGCTAGTCTGGCAGCGCCCGCTTTCCTCTCTCTCT 120
61 AAGTGTTCACAGGCTACCCGCTAGTCTGGCAGCGCCCGCTTTCCTCTCTCTCT 120
121 CGCGTGGCGGGGAACTGTTGGCGCGCGGCTTCGGGAGCGCCAGCTCCCGCGCG 180
121 CGCGTGGCGGGGAACTGTTGGCGCGCGGCTTCGGGAGCGCCAGCTCCCGCGCG 180
121 CGCGTGGCGGGGAACTGTTGGCGCGCGGCTTCGGGAGCGCCAGCTCCCGCGCG 180
181 AGGTCCCGGCGAGTAAACATAGATCATCATAGAGAACTCTTGAAGTTGTTCAAGAAA 240
181 AGGTCCCGGCGAGTAAACATAGATCATCATAGAGAACTCTTGAAGTTGTTCAAGAAA 240
241 ATTGAAAGTACGCAAAATAGAAATTAAGAAATTAACAGACATACAGACAGCATGGA 300
241 ATTGAAAGTACGCAAAATAGAAATTAAGAAATTAACAGACATACAGACAGCATGGA 300
301 AGTGTGCTTAGGAAACAGAACAGACAGAGGAAAGAAATCCGCTCGATAC 360
301 AGTGTGCTTAGGAAACAGAACAGACAGAGGAAAGAAATCCGCTCGATAC 360
301 AGTGTGCTTAGGAAACAGAACAGACAGAGGAAAGAAATCCGCTCGATAC 360
361 AACTGACAGTATATGTTTCCGCGCTTCAATGCTTTAGATGGATGCTCTTTGTCA 420
361 AACTGACAGTATATGTTTCCGCGCTTCAATGCTTTAGATGGATGCTCTTTGTCA 420
361 AACTGACAGTATATGTTTCCGCGCTTCAATGCTTTAGATGGATGCTCTTTGTCA 420
421 TAAAGTGCATTTTACATGCGCAACAGTAACTCTTACCGAAGTGTGCTCAGAAAT 480
421 TAAAGTGCATTTTACATGCGCAACAGTAACTCTTACCGAAGTGTGCTCAGAAAT 480
481 ATTTAGTACATTTGCAACAGGCTCTGGAAGAACTGAAGAACTGAAGAACTGAAG 540
481 ATTTAGTACATTTGCAACAGGCTCTGGAAGAACTGAAGAACTGAAGAACTGAAG 540
481 ATTTAGTACATTTGCAACAGGCTCTGGAAGAACTGAAGAACTGAAGAACTGAAG 540
541 ACTATGAAATTCAGTTGCGCAAGGTTAATGTTGTTCAAGAAATATCAAGATACCTG 600
541 ACTATGAAATTCAGTTGCGCAAGGTTAATGTTGTTCAAGAAATATCAAGATACCTG 600
601 GAAA 604
601 GAAA 604

RESULT 8
B1712516/c
LOCUS
DEFINITION B1712516 593 bp mRNA linear EST 11-MAR-2002
1e08h05.X1 H85 islet Homo sapiens CDNA clone IMAGE:5086689 3',
mRNA sequence.
ACCESSION B1712516
VERSION B1712516.1 GI:15688211
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Mammalia; Euthera; Primates; Carnivora; Homidae; Homo.
1 (bases 1 to 593)
Mellon, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Keastner, K.,
Lemishka, I., Scaer, M., Brestelli, J., Gradow, G., Clifton, S.,
Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Bistul, A.,
Schmitt, A., Theising, B., Rittler, E., Ronko, I., Bennett, J., Cardenas,
M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.,
Jackson, Y., and Bowers, Y.

TITLE
JOURNAL
COMMENT
Unpublished (2000)
Contact: Douglas Mellon, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmellon@iobp.harvard.edu

Laboratory was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)
Seq primer: -400p from Glibco
High quality sequence stop: 430.
Location/Qualifiers
1. 593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5086689"
/clone_lib="H85 islet"
/issue_type="Purified pancreatic islet"
/lab_host="DH10B"

FEATURES
SOURCE

/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; CDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."

BASE COUNT 237 a 93 c 91 g 172 t
ORIGIN

Query Match 12.9%; Score 583.6; DB 13; Length 593;
Best Local Similarity 99.0%; Pred. No. 1.4e-100;
Matches 586; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

3916 CATTCCTGAGGAGGAGAAAGTATGCAAGAGTAAATTTTCATTCCTCAATCACTAGC 3975
3916 CATTCCTGAGGAGGAGGAGAAAGTATGCAAGAGTAAATTTTCATTCCTCAATCACTAGC 3975
593 CATTCCTGAGGAGGAGGAGAAAGTATGCAAGAGTAAATTTTCATTCCTCAATCACTAGC 534
3976 TGTAGACTGATCTGTTGACAGCTGTTGCTCATTTTGTCTGCTGCTGCTGCTGCTG 4035
3976 TGTAGACTGATCTGTTGACAGCTGTTGCTCATTTTGTCTGCTGCTGCTGCTGCTG 4035
533 TGTAGACTGATCTGTTGACAGCTGTTGCTCATTTTGTCTGCTGCTGCTGCTGCTG 474
4036 GACATTTGTTGAGATATTCATTTGCTGCTCATGATATTTTCTTTTAATATCACT 4095
4036 GACATTTGTTGAGATATTCATTTGCTGCTCATGATATTTTCTTTTAATATCACT 4095
473 GACATTTGTTGAGATATTCATTTGCTGCTCATGATATTTTCTTTTAATATCACT 414
4096 TGATATCTTGTCTTAATTTTCTTACATATGTTTGCCTGATACAGATTTTAT 4155

```

Db 413 TGAATACCTGTTCTTAAATTTCTTCACATATGGTTGGCTGATCAACTGATTTTAT 354
OY 4156 AACTGAAATTTAAGATCATCAGCTAAACACAGTAAGTGCATTTATTTCCCTTTAAC 4215
Db 353 AACTGAAATTTAAGATCATCAGCTAAACACAGTAAGTGCATTTATTTCCCTTTAAC 294
OY 4216 ATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTTTTCCTGAGCATGTGAG 4275
Db 293 ATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTTTTCCTGAGCATGTGAG 244
OY 4276 CCGATTAATACCATTTTATTTGCTTTATTTCTAATATGGAACAATGAGTGAAT 4335
Db 233 CCGATTAATACCATTTTATTTGCTTTATTTCTAATATGGAACAATGAGTGAAT 174
OY 4336 CTAATATATGTTGTAGTATATAAACATCATTTAGCCTAATTTATAGAAATGCTAATTA 4395
Db 173 CTAATATATGTTGTAGTATATAAACATCATTTAGCCTAATTTATAGAAATGCTAATTA 114
OY 4396 GTACACGACACATAGAAACATGAATTTGCTTATGATTTGACCTTTGTCAGCAATTTTGA 4455
Db 113 GTGCGACGACATAGAAACATGAATTTGCTTATGATTTGACCTTTGTCAGCAATTTTGA 54
OY 4456 AGTCATTAATGTTGTATCATTAATTTAATTAAGTGTGCTGGCTTTCAGAAATAC 4507
Db 53 AGTCATTAATGTTGTATCATTAATTTAATTAAGTGTGCTGGCTTTCAGAAATAC 2

RESULT 9
AM961708 612 bp mRNA linear EST 01-JUN-2000
LOCUS EST373781 MAGE resequences, MAGG Homo sapiens cDNA, mRNA sequence.
ACCESSION AM961708
VERSION AM961708.1 GI:8151392
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 612)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
JOURNAL The Institute for Genomic Research
COMMENT 9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@ligr.org
Plate: 164
Seq primer: Reverse.
FEATURES
Source Location/Qualifiers
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGG"
/note="Vector: pBluescriptSKm"
BASE COUNT 202 a 119 c 116 g 173 t 2 others
ORIGIN
Query Match 12.8%; Score 581.2; DB 10; Length 612;
Best Local Similarity 98.9%; Pred. No. 4e-100;
Matches 605; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
OY 1912 GACTATTTAGTCCACCATGAAACAGCAAAAGAGATTTTGTAGAGCAGAACTACC 1971
Db 1 GACTATTTAGTCCACCATGAAACAGCAAAAGAGATTTTGTAGAGCAGAACTACC 60
OY 1972 TAAAGGATATGTTATCACTGGAATTTATTTGAGACAGATGTTTGTACTGTCAACA 2031

```

```

Db 61 TAAAGGATATGTTATCACTGGAATTTATTTGAGAGAGATGTTTGTACTGTCAACA 120
OY 2032 AATATGCTGCAGATCTTCCAGCCCTGCTGCTCCAGACACAGAGCAAAATAGAGA 2091
Db 121 AATATGCTGCAGATCTTCCAGCCCTGCTGCTCCAGACACAGAGCAAAATAGAGA 180
OY 2092 GCATCCCACTAGCTAGACACATGACACAAGACATAGTCAATATTAACAGATGACTAC 2151
Db 181 GCATCCCACTAGCTAGACACATGACACAAGACATAGTCAATATTAACAGATGACTAC 240
OY 2152 TGGAAATGTTTCCGGAATCACTGTGAAATGCTCCGATTTTTCAGACTTCAGAAAC 2211
Db 241 TGGAAATGTTTCCGGAATCACTGTGAAATGCTCCGATTTTTCAGACTTCAGAAAC 300
OY 2212 CATTATTTATTTTGTTCAGTATGACGACGTGTAATCTCAATATTAAGCAATTTGA 2271
Db 301 CATTATTTATTTTGTTCAGTATGACGACGTGTAATCTCAATATTAAGCAATTTGA 360
OY 2272 CACTGTTAAAGCAGAAATACCTTGATTCATTACTCAGTCTGTTTAAATCTTAAGAAATA 2331
Db 361 CACTGTTAAAGCAGAAATACCTTGATTCATTACTCAGTCTGTTTAAATCTTAAGAAATA 420
OY 2332 CTCACGTGGGAGAGGAATCTTGGGCAATTTTATGTCCTGCTCCCTGCTTTC 2391
Db 421 CTCACGTGGGAGAGGAATCTTGGGCAATTTTATGTCCTGCTTTCCTTCTTTC 480
OY 2392 TTGTTTGTGTAATCTGCAATTCAGTGTGCGCAAGTATTTGCTTCCCTGAGACAGCTTA 2451
Db 481 TTGTTTGTGTAATCTGCAATTCAGTGTGCGCAAGTATTTGCTTCCCTGAGACAGCTTA 540
OY 2452 TAATTAAGAAACCTTGTATTTGTGCTGAG- AAATTGAGACAGGAC-TAGAAATCA 2509
Db 541 TAATTAAGAAACCTTGTATTTGTGCTGAG- AAATTGAGACAGGAC-TAGAAATCA 600
OY 2510 TATCACAATTTT 2521
Db 601 TATCACAATTTT 612

RESULT 10
BG928131 686 bp mRNA linear EST 06-NOV-2001
LOCUS BG928131
DEFINITION HNC65-1-B8.R.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
sequence.
ACCESSION BG928131
VERSION BG928131.1 GI:14322654
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 686)
Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
Lark,M.W.
Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthr. Cartil. 9 (7), 641-653 (2001)
JOURNAL 21482651
MEDLINE Contact: Sanjay Kumar
COMMENT UW2109
GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-1@sk.com
Seq primer: 17.
FEATURES
Source Location/Qualifiers
1..686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HNC (Human Normal Cartilage)"

```

/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"

BASE COUNT 188 a 132 c 127 g 239 t
ORIGIN

Query Match 12.6%; Score 568.2; DB 13; Length 686;
Best Local Similarity 99.3%; Pred. No. 1.1e-97;
Matches 581; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 3589 CATGTTTGAAAAATTTGGCTATTAAAGATATGATTAGATGGCTTATCTGATTTAT
DB 86 CATGTTTGAAAAATTTGGCTATTAAAGATATGATTAGATGGCTTATCTGATTTAT
QY 3649 ACCGAGTACACCTGATCTTTCTATATTTCAGAAAGTATGGGATTAACCTAGAG
DB 146 ACCGAGTACACCTGATCTTTCTATATTTCAGAAAGTATGGGATTAACCTAGAG
QY 3709 AGAGCTAGATGATTTATTTTAAAGTATGATCTTTAAACCTCTTATTTCTACAA
DB 206 AGAGCTAGATGATTTATTTTAAAGTATGATCTTTAAACCTCTTATTTCTACAA
QY 3769 GTTATATGGCTAAATTCAGATGACAGGATTCAGCATTCCTCATGAGAA
DB 266 GTTATATGGCTAAATTCAGATGACAGGATTCAGCATTCCTCATGAGAA
QY 3829 AGAGAGCTCCCTCATCTGAAAGCTCTGAAATCTACCTTCGACAGCTTACAGCAATC
DB 326 AGAGAGCTCCCTCATCTGAAAGCTCTGAAATCTACCTTCGACAGCTTACAGCAATC
QY 3889 AGTATGATCTCCCTGAGCAGCAGGCTCATCTGAGAGGAGGAAAGATTAGCCAAAGA
DB 386 AGTATGATCTCCCTGAGCAGCAGGCTCATCTGAGAGGAGGAAAGATTAGCCAAAGA
QY 3949 GTTATATTTTCATCCAAATCACTTACCTGTTAGACTGATCTGTTTGTAGCAGTTGTTGT
DB 446 GTTATATTTTCATCCAAATCACTTACCTGTTAGACTGATCTGTTTGTAGCAGTTGTTGT
QY 4009 CTCATTTTGTCTGTCGATTTTGTGAGACATTTGTGAGAAATATCTTATTTGGTCTCT
DB 506 CTCATTTTGTCTGTCGATTTTGTGAGACATTTGTGAGAAATATCTTATTTGGTCTCT
QY 4069 ACTGATATTTTCTTTTAAATCTACTGATGATCTGTTTAAATTTCTTACATAT
DB 566 ACTGATATTTTCTTTTAAATCTACTGATGATCTGTTTAAATTTCTTACATAT
QY 4129 GG-TTGGCTGATACAACTGATTTTAACTGAATTTAAGAA 4172
DB 626 GGTTTGGCTGATACAACTGATTTTAACTGAATTTAAGGA 670

RESULT 11
BM972493/c 585 bp mRNA linear EST 21-MAR-2002

LOCUS
DEFINITION
UI-CF-EC1-abq-p-23-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone

ACCESSION
BM972493
VERSION
BM972493.1 GI:19590079

KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 585)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL
Genome Res. 6 (9), 791-806 (1996)

MEDLINE
97044477

COMMENT
Contact: McCray, PB
McCrory Lab
University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul.mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLY-A=yes.

FEATURES
SOURCE
Location/Qualifiers

1..585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-EC1-abq-p-23-0-UI"
/clone_1db="UI-CF-EC1"
/tissue_type="Lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pRT3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pRT3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dr)18 tail. The sequence tag for this library is
AAGGCTTAC.
TAG_LIB=UI-CF-EC1
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG_SEQ=AAAGTCTTAC"

BASE COUNT 228 a 83 c 92 g 182 t

ORIGIN

Query Match 12.5%; Score 567; DB 14; Length 585;
Best Local Similarity 97.9%; Pred. No. 2e-97;
Matches 573; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 3933 AAAGATTAGCCAAAGATTAAATTTTCATTCCAATCACTGATGATGATCTGTT 3992
DB 585 AAAGATTAGCCAAAGATTAAATTTTCATTCCAATCACTGATGATGATCTGTT 526
QY 3993 TGTAGCAGTTGTTGTCATTTTGTGTCGTGCAATTTTGGAGCATTGTTGAGAAATA 4052
DB 525 TGTAGCAGTTGTTGTCATTTTGTGTCGTGCAATTTTGGAGCATTGTTGAGAAATA 466
QY 4053 TTCTATTTGGGTCCTACTGATTTTCTTTTAAATATCTACTGATATCTTTGTTCTTTA 4112
DB 465 TTCTATTTGGGTCCTACTGATTTTCTTTTAAATATCTACTGATATCTTTGTTCTTTA 406
QY 4113 AATTTTCTTCAATATGTTGGCTGATACAACTGATTTTAACTGAATTTAAGGAA 4172
DB 405 AATTTTCTTCAATATGTTGGCTGATACAACTGATTTTAACTGAATTTAAGGAA 346
QY 4173 TCTAAGACCTAAACTGATGATGATATTTCTTAAATACATAGACCGTGTCTACT 4232
DB 345 TCTAAGACCTAAACTGATGATGATATTTCTTAAATACATAGACCGTGTCTACT 286
QY 4233 CTCAGACACCTCTCTCATATTTTCTCTGATGATGATGATGATGATGATGATGATGAT 4292
DB 285 CTCAGACACCTCTCTCATATTTTCTCTGATGATGATGATGATGATGATGATGATGAT 226
QY 4293 TCATTTGCTTTTATTTTAAATATGGAACAATGAGATGAACTGATTAATATAGTTGATG 4352


```
Db 225 TCATTGCTTTTATTTCTTAATATGGAACATGAGTGAAGTCTTAATATAGCTGTGAG 166
QY 4353 TATATAAATCATGATAGCCATTAATTTAGTAAATGCTAATTTAGTACGACACATAGAAA 4412
Db 165 TATATAAATCATGATAGCCATTAATTTAGTAAATGCTAATTTAGTACGACACATAGAAA 106
QY 4413 CATGAATTTGCTTACATTTGACCTTTTGTGACCAATTTTGTACAGTCATTAATGTTTGTTC 4472
Db 105 CATGAATTTGCTTACATTTGACCTTTTGTGACCAATTTTGTACAGTCATTAATGTTTGTTC 46
QY 4473 ATATATTTTAAATTAAGTGTCTGGGTTTCAAGTAAATCTTCAAAAAA 4517
Db 45 ATATATTTTAAATTAAGTGTCTGGTTCAGAAAAA 1

RESULT 12
AL040840/c 564 bp mRNA linear EST 29-FEB-2000
LOCUS DKEP434F0915_s1 434 (synonym: htes3) Homo sapiens cDNA clone
DEFINITION DKEP434F0915 3', mRNA sequence.
ACCESSION AL040840
VERSION AL040840.1 GI:5409785
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 564)
AUTHORS Bloeker,H., Boecker,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
TITLE EST (Bioecker, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Bloecker H
MIPS
Am Klopfersplitz 18a D-82152 Martinsried, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Blaunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No r1 sequence available.
This clone (DKEP434F0915) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
location/Qualifiers
source 1..564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKEP434F0915"
/clone_lib="434 (synonym: htes3)"
/tissue.type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 221 a 88 c 95 g 160 t
ORIGIN
Query Match 12.5%: Score 563.6; DB 9: Length 564;
Best Local Similarity 99.8%; Pred. No. 8.8e-97;
Matches 563: Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4090 TCTACTGTATATCTGTCTTTAAATTTTCTTCACATATGTTGGCTGATACACTGAT 4149
Db 384 TCTACTGTATATCTGTCTTTAAATTTTCTTCACATATGTTGGCTGATACACTGAT 325
QY 4150 TTTTATACTGAATTTAAGATCTTACACGCTAAACACAGTAAGTATGATTTTCTCT 4209
Db 324 TTTTATACTGAATTTAAGATCTTACACGCTAAACACAGTAAGTATGATTTTCTCT 265
QY 4210 TATACATAGACCCGTTGTCTACTCTCAGACCCCTCTCCAAATTTTTCCTGTAGCAT 4269
Db 264 TATACATAGACCCGTTGTCTACTCTCAGACCCCTCTCCAAATTTTTCCTGTAGCAT 205
QY 4270 GTGATGCCGTGATTTAAACGATTTTCAATTTGCTTTATTTCTAATATGGAACATGAGAG 4329
Db 204 GTGATGCCGTGATTTAAACGATTTTCAATTTGCTTTATTTCTAATATGGAACATGAGAG 145
QY 4330 TGAACCTTAATATAGTGTAGTATATAAACAATCTTGGCCATTAATTTAGAAATGCT 4389
Db 144 TGAACCTTAATATAGTGTAGTATATAAACAATCTTGGCCATTAATTTAGAAATGCT 85
QY 4390 AATTAGTACACGACATGAAACATGAAATGCTTACATTTGATCTTGTGACCAAT 4449
Db 84 AATTAGTACACGACATGAAACATGAAATGCTTACATTTGATCTTGTGACCAAT 25
QY 4450 TTTGACAGTCATTAATGTTTGTCA 4473
Db 24 TTTGACAGTCATTAATGTTTGTCA 1

RESULT 13
BE895218 900 bp mRNA linear EST 20-OCT-2000
LOCUS 60143654F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:391887 5',
DEFINITION mRNA sequence.
ACCESSION BE895218.1 GI:10358391
VERSION BE895218
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 900)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs.rem@nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9747 row: 1 column: 14
High quality sequence scop: 557.
location/Qualifiers
source 1..900
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:391887"
/clone_lib="NIH_MGC_72"
/tissue.type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 252 a 155 c 177 g 315 t 1 others
ORIGIN
Query Match 12.4%: Score 562; DB 12: Length 900;
```


Db 598 TCAATTATGGCTGGCTGATA-AACGTGATTCTATACTGGAATTCAGGAT-TAACAG 655
 QY 4181 CTAAACTCAGTAAGTGC 4198
 Db 656 CTAAACTCAGTAGGGC 673

RESULT 15 AM65394/c

LOCUS hi90C01.x1 Soares_NFL.T.GBC.S1 Homo sapiens mRNA linear EST 06-APR-2000
 DEFINITION IMAGE:2979552 3', mRNA sequence.

ACCESSION AM65394
 VERSION AM65394.1 GI:7457940

KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 556)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40UP from Gibco

FEATURES
 Location/Qualifiers
 1..556

Source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2979552"
 /clone_lib="Soares_NFL.T.GBC.S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT7T3P-Pac (Pharmacia) with
 a modified polylinker; Site: 1: Not 1; Site: 2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI CGAP GCBI) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 224 a 80 c 90 g 161 t 1 others
 ORIGIN

Query Match 12.2%; Score 552.6; DB 10; Length 556;
 Best Local Similarity 99.6%; Pred. No. 1,1e-94;
 Matches 552; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3953 ATTTTCATCCAAATCATTAGCTGTAGCTGTTGTAGCAGTTGTTGCTCA 4012
 Db 556 ATTTTCATCCAAATCATTAGCTGTAGCTGTTGTAGCAGTTGTTGCTCA 497
 QY 4013 TTTTGTCTGTCGATTTTGTAGACATTTGTGAGAAATTTGTTGCTCACTG 4072
 Db 496 TTTTGTCTGTCGATTTTGTAGACATTTGTGAGAAATTTGTTGCTCACTG 437
 QY 4073 TATTTTCTTTTAATATCTACTGATATCTGTTTAAATTTTCTTCACATATGTT 4132
 Db 436 TATTTTCTTTTAATATCTACTGATATCTGTTTAAATTTTCTTCACATATGTT 377
 QY 4133 TGGCTGATACACTGATTTTATTAAGAAATTAAGAACTAAAGCTCAAGT 4192
 Db 376 TGGCTGATACACTGATTTTATTAAGAAATTAAGAACTAAAGCTCAAGT 317
 QY 4193 AAGTGATATATTTCTTATTAACATAGACCCGTTGCTACTCTCAGACCCCTCTCAAT 4252

Db 316 AAGTGATATATTTCTTATTAACATAGACCCGTTGCTACTCTCAGACCCCTCTCAAT 257
 QY 4253 TTTTTCCTGTAGCATGTGATGCTGATTAACTCATTTTGCCTTTTATTTCTAA 4312
 Db 256 TTTTTCCTGTAGCATGTGATGCTGATTAACTCATTTTGCCTTTTATTTCTAA 197
 QY 4313 TATGGACAATAGAGAGTGAAGTCTTAATATATAGTTGTAGTAATTAACATCATATAGCTT 4372
 Db 196 TATGGACAATAGAGAGTGAAGTCTTAATATATAGTTGTAGTAATTAACATCATATAGCTT 137
 QY 4373 AATTATTAAGAAATGCTAATTAGTACCACATAGAAACATGAATGCTTAGTCATT 4432
 Db 136 AATTATTAAGAAATGCTAATTAGTACCACATAGAAACATGAATGCTTAGTCATT 77
 QY 4433 GTACCTTTGCACCAATTTTGCACATCAATTAATGTTGCTATATTTTAAATTAAGTGC 4492
 Db 76 GTACCTTTGCACCAATTTTGCACATCAATTAATGTTGCTATATTTTAAATTAAGTGC 17
 QY 4493 TGGGTTTCAGAAATA 4506
 Db 16 TGGGTTTCAGAAATA 3

Search completed: March 22, 2003, 10:47:45
 Job time: 6639 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:10:12 ; Search time 39 Seconds
(without alignments)
2757.263 Million cell updates/sec

Title: US-09-847-046-2
Perfect score: 4142
Sequence: 1 MFSGFNFRVQISFVIMCIF.....FKAEKSPRDKELGCSKVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

```
1: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/genescp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/genescp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/genescp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/genescp-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4142	100.0	807	ABG61812	Prostate cancer-as
2	1839	44.4	357	AA894073	Human protein sequ
3	185	4.5	485	AA849829	Highly heat-resist
4	185	4.5	505	AA873963	Humicola insolens
5	180	4.3	505	AA863807	Protein disulphide
6	177	4.3	496	AB862855	Drosophila melanog
7	175.5	4.2	505	AA883801	Bovine phospholipa
8	172.5	4.2	511	AA865506	Aspergillus sp. re
9	172.5	4.2	515	AA865504	Aspergillus sp. re
10	172.5	4.2	515	AA849946	A. oryzae protein

11	169.5	4.1	515	22	AAE12991
12	169.5	4.1	515	22	AA882539
13	169.5	4.1	544	21	AA843424
14	169.5	4.1	549	23	ABP41717
15	166.5	4.0	505	16	AA870329
16	166.5	4.0	515	16	AA869505
17	164	4.0	454	22	AA849336
18	155	3.7	450	22	AA849944
19	153.5	3.7	489	21	AAV70766
20	153.5	3.7	526	13	AA826174
21	150	3.6	579	21	AA841949
22	148.5	3.6	491	21	AAV70770
23	146.5	3.5	492	23	AB807836
24	145	3.5	918	22	AAU30491
25	141.5	3.4	491	21	AAV70769
26	139	3.4	496	21	AAV70762
27	138.5	3.3	416	22	AB858589
28	138.5	3.3	645	22	AA866531
29	138.5	3.3	646	22	AAU30243
30	138	3.3	522	19	AA869726
31	136	3.3	466	23	ABG60996
32	136	3.3	470	23	ABG61002
33	136	3.3	495	19	AAW70270
34	136	3.3	495	23	ABG61000
35	136	3.3	1579	23	AB891088
36	135.5	3.3	510	9	AA80615
37	135	3.3	460	21	AA841950
38	135	3.3	468	23	AA866107
39	135	3.3	472	23	AA866113
40	135	3.3	491	13	AA825296
41	135	3.3	497	23	AA866111
42	135	3.3	508	9	AA80664
43	135	3.3	508	13	AA825297
44	135	3.3	508	15	AA851696
45	135	3.3	513	22	AA865635

ALIGNMENTS

RESULT 1	
ABG61812	
ID	ABG61812 standard; Protein; 807 AA.
AC	ABG61812;
DT	15-AUG-2002 (first entry)
DE	Prostate cancer-associated protein #13.
KW	Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
OS	Mammalia.
XX	
PN	WO200230268-A2.
XX	
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US32045.
XX	
PR	13-OCT-2000; 2000US-0687576.
XX	
PR	08-DEC-2000; 2000US-0733288.
XX	
PR	08-DEC-2000; 2000US-0733742.
XX	
PR	24-JAN-2001; 2001US-2639576.
XX	
PR	16-MAR-2001; 2001US-276791P.
XX	
PR	06-MAR-2001; 2001US-276688P.
XX	
PR	24-APR-2001; 2001US-281922P.
XX	
PR	30-APR-2001; 2001US-286214P.
XX	
PR	04-MAY-2001; 2001US-288589P.
XX	
PA	(EOSB) EOS BIOTECHNOLOGY INC.
XX	

Human protein disu
Human protein disu
Human cancer assoc
Human ovarian anti
Human phospholipas
Aspergillus sp. re
Protein disulfide
Protein disulfide
Spider protein dis
Protein disulphide
Arabidopsis thalia
Scorpion protein d
Human protein iden
Novel human secret
Scorpion protein d
Spider protein dis
Drosophila melanog
Human interferon-a
Novel human secret
Saccharomyces cere
Dirofilaria immiti
Dirofilaria immiti
Dirofilaria immiti
Herbicidally activ
Bovine protein dis
Arabidopsis thalia
D. immitis transgl
Recombinant PDI (A
D. immitis transgl
Polypeptide with P
PDI. Homo sapiens
Human PDI. Homo s
Amino acid sequenc

PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
 XX WPI; 2002-471335/50.
 DR N-PSDB; ABK92127.
 XX
 PT Detecting a prostate cancer-associated transcript in a cell in a
 PT patient, useful for diagnosing prostate cancer (PC) or screening
 PT modulators of PC, by determining if prostate cancer-associated genes
 PT are expressed in a prostate tissue -
 XX
 PS Claim 27, Page 311; 436pp; English.
 CC
 CC The present invention relates to methods of detecting a prostate
 CC cancer-associated transcript in a cell from a patient. The method
 CC comprises contacting a biological sample from the patient with
 CC prostate cancer-associated polynucleotides (designated PC genes) that
 CC selectively hybridise to a sequence that is at least 80% identical
 CC to them. The prostate cancer-associated polynucleotide sequences
 CC are differentially expressed in prostate tumour tissue or in
 CC prostate cancer and are derived from the tissues of various
 CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
 CC The methods of the invention are useful for diagnosing and treating
 CC prostate cancer in mammals. The prostate cancer-associated genes are
 CC useful for diagnosing or treating prostate cancer, as well as for
 CC identifying modulators of prostate cancer or agents that inhibit
 CC prostate cancer. The nucleic acid sequences are particularly useful
 CC in gene therapy, as a vaccine or in antisense applications.
 CC ABG61800-ABG61944 represent prostate cancer-associated proteins.
 CC
 XX
 SQ Sequence 807 AA;
 Query Match 100.0%; Score 4142; DB 23; Length 807;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 807; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 1 MESSFNFRGIGSPVIMCIFYMPVNSLPESLPQKVFSTQPGLELNENRPRQDYGIS 60
 DB 1 MESSFNFRGIGSPVIMCIFYMPVNSLPESLPQKVFSTQPGLELNENRPRQDYGIS 60
 QY 61 VAKNCVKEEISRYCGKEKDKMAVLEFGNLLREFPTDLEFDNAIVAHVLFALLSEV 120
 DB 61 VAKNCVKEEISRYCGKEKDKMAVLEFGNLLREFPTDLEFDNAIVAHVLFALLSEV 120
 QY 121 KYTNLEDLONENALGKANIIFSYVRAIGIPENRAVMEAGVYGTYYQFVLTETALL 180
 DB 121 KYTNLEDLONENALGKANIIFSYVRAIGIPENRAVMEAGVYGTYYQFVLTETALL 180
 QY 181 ESTGSEDEVEXAHLVFEHCKIVLDTLOCCRTLMROPITTTINHLFITMKARPLTVEAD 240
 DB 181 ESTGSEDEVEXAHLVFEHCKIVLDTLOCCRTLMROPITTTINHLFITMKARPLTVEAD 240
 QY 241 POOVSTVHLQGLPLVFIVSQATYEADRTAEVAVWRLLGKAGVLLLRDSLEVNTPOD 300
 DB 241 POOVSTVHLQGLPLVFIVSQATYEADRTAEVAVWRLLGKAGVLLLRDSLEVNTPOD 300
 QY 301 ANVFKRAEGGVPEFLVLDVLLISHVENNHIEIOEDEDDMEGPDIDVDDEVAE 360
 DB 301 ANVFKRAEGGVPEFLVLDVLLISHVENNHIEIOEDEDDMEGPDIDVDDEVAE 360
 QY 361 TVPRDRRKRLPLELTVLLEETENATYMASDSYLFYAGQAVSMAPLOSTIDVAVYKLG 420
 DB 361 TVPRDRRKRLPLELTVLLEETENATYMASDSYLFYAGQAVSMAPLOSTIDVAVYKLG 420
 QY 421 TSTMLLRINACADMSDVCYQNTVEFPIIKMYKKGENDPVSYAGMLGTRDKLFTQLNRIS 480
 DB 421 TSTMLLRINACADMSDVCYQNTVEFPIIKMYKKGENDPVSYAGMLGTRDKLFTQLNRIS 480
 QY 481 YPVNITSIOGAEEYLGELYKDLLVSSVYGLFSPMTKANEDFSEAGNYLKGVIYTG 540
 DB 481 YPVNITSIOGAEEYLGELYKDLLVSSVYGLFSPMTKANEDFSEAGNYLKGVIYTG 540
 QY 541 TYSEEDVLLSTKYAASLPLALLARHTEGKIESIPLASTAODIVQIITDALLEMPEIT 600
 DB 541 TYSEEDVLLSTKYAASLPLALLARHTEGKIESIPLASTAODIVQIITDALLEMPEIT 600

DB 541 TYSEEDVLLSTKYAASLPLALLARHTEGKIESIPLASTAODIVQIITDALLEMPEIT 600
 QY 601 VENLPSEYERLQKPLLIIESDGTVPNOYKKAILLTVKOKYLDSPFCMLKNPVPVGGIL 660
 DB 601 VENLPSEYERLQKPLLIIESDGTVPNOYKKAILLTVKOKYLDSPFCMLKNPVPVGGIL 660
 QY 661 RAYEDPLPPLVLVNLHSGGVFAFPSPDAILIENMLVLMKLEAGLENHITILPAQE 720
 DB 661 RAYEDPLPPLVLVNLHSGGVFAFPSPDAILIENMLVLMKLEAGLENHITILPAQE 720
 QY 721 WKPLPAYDPLSMIDATSGRTKRVKCKMETDVOENDKQEHDKSAVREPLETLRIK 780
 DB 721 WKPLPAYDPLSMIDATSGRTKRVKCKMETDVOENDKQEHDKSAVREPLETLRIK 780
 QY 781 HMNRSNMFKEAKESFRDKELGCSKVN 807
 DB 781 HMNRSNMFKEAKESFRDKELGCSKVN 807
 RESULT 2
 AAB94073
 ID AAB94073 standard; Protein; 357 AA.
 XX
 AC AAB94073;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:14262.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PE 28-JUL-2000; 2000EP-0116126.
 PF
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000CP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 14262; 2537bp + CD ROM; English.
 CC
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH13632
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 357 AA;

Query Match 44.4%; Score 1839; DB 22; Length 357;
 Best Local Similarity 99.7%; Pred. No. 2.6e-155;
 Matches 356; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 451 MYKGENPVSAAGMGTDLKFTQNLNRSYVNTSIOEAEVYSGELKDLITYSSVS 510
 Db 1 MYKGENPVSAAGMGTDLKFTQNLNRSYVNTSIOEAEVYSGELKDLITYSSVS 60
 QY 511 VLGLFSPMTAKEDFSEAGNYLKGVTGTYSEEDVLLSPKYASLPALLARHTEGK 570
 Db 61 VLGLFSPMTAKEDFSEAGNYLKGVTGTYSEEDVLLSPKYASLPALLARHTEGK 120
 QY 571 IESIPLASTHAODIVQITTDALLEMFEPIYENLPSYRLOKPLILFSDGTVPNQYKA 630
 Db 121 IESIPLASTHAODIVQITTDALLEMFEPIYENLPSYRLOKPLILFSDGTVPNQYKA 180
 QY 631 ILTLVKQKYLDSFTPCWNLKNTKTPYRGILRAYFPPLPLVLYVNLHSGQVAFPSD 690
 Db 181 ILTLVKQKYLDSFTPCWNLKNTKTPYRGILRAYFPPLPLVLYVNLHSGQVAFPSD 240
 QY 691 QAIIEENLVMLKLEAGLENNHITLPAQEWKPLPAYDFLSMIDATSORGTTRVPCW 750
 Db 241 QAIIEENLVMLKLEAGLENNHITLPAQEWKPLPAYDFLSMIDATSORGTTRVPCW 300
 QY 751 KETDVOENDKQHEKDSAVRKEPIETLRKIHNNRSNMEKEKSPRRKELGCSKVN 807
 Db 301 KETDVOENDKQHEKDSAVRKEPIETLRKIHNNRSNMEKEKSPRRKELGCSKVN 357

RESULT 3

AAH9829 ID AAR49829 standard; protein; 485 AA.

XX AAR49829;

DT 29-SEP-1994 (first entry)

XX Highly heat-resistant protein disulphide isomerase.

KW Protein disulphide isomerase; PDI; 3-D; heat; resistant;
 catalysis; enzyme; ribonuclease A; substrate; refolding.

XX Humicola insolens KASI.

PN JP06038752-A.

PD 15-FEB-1994.

XX 04-MAR-1993; 93JP-0044013.

XX 27-MAY-1992; 92JP-0135254.

XX (TOYAMA) TOYAMA CHEM CO LTD.

XX WPI; 1994-094827/12.

PT Novel protein disulphide isomerase active substance - for use in
 the production of 3-D active proteins

PS Claim 2; Page 7-8; 10pp; Japanese.

XX The new PDI has the following characteristics: (A) has an ability of

CC catalysing disulphide exchange reaction in a protein; (B) uses at
 CC least ribonuclease A as the substrate; (C) has an optimum acting
 CC temp. of 20 to 70 degrees C; (D) is stable at pH 6 to 9; and (E) has a
 CC mol.wt. of ca. 60000 to 62000 measured by SDS-PAGE. PDI can be used
 CC advantageously in a refolding reaction of a protein.

XX Sequence 485 AA;

Query Match 4.5%; Score 185; DB 15; Length 485;
 Best Local Similarity 21.9%; Pred. No. 1.9e-07;
 Matches 112; Conservative 79; Mismatches 213; Indels 108; Gaps 22;

QY 15 VIMCIYFMYTVNSLDELSPQKYSTLOPGLNEAVRLODYGISAVKNCVKEISRY 74
 Db 20 LVLAEPFAPWCHCKALAP-----EYEEATTLKKNKIKAKVDT----- 60
 QY 75 CGKEKDKAKYLFKGNILREF-----PTTLEDVNAVYAVYFALL--FSEVKYITN 125
 Db 61 --EETDLCOQHGVCEYPTLKFVRGLDNVSPYQGRKAAITSYMIKOSLPAYSEVTK-DN 117
 QY 126 LEDLQNIEMALKGANIIFSYVRAIGIPHRVMEAGFYGTTOFVLTTETALLSTGS 185
 Db 118 LEEFKKADKA-----VLVAYVDASDKASSEVFTQVAEKLRDNPFGSSSDALAEBSV 171
 QY 186 EDVEYAHLYFFHCKLVLDLTQCCRTIMEOPLTITNHLFTMKAPLLEVAEDPOOVS 245
 Db 172 KAP-----ATVLKDFDEGKAVSEKREVEPAIEKFAATGATPLIGETG--PEYTS 219
 QY 246 TVHLOGLPLRYFVSQA-----TYEADRTAEW--VAMRLIG--NAGVLLLR 290
 Db 220 D-YMSAGIPLAIFYFAETAEERKELSDKLPIEAORGVINFTIDAKAFGAHAGNINLKT 278
 QY 291 DSLEY-----NIPQDANVVFKRAEGVPYFVLVHVDLISIVENNMHTEQDEEDNDM 346
 Db 279 DKFPAFAIOEYAKNKPFPDQKE--TFEPAIAEYVDDEPAGKIEPISSEPIPEKO---- 333
 QY 347 EGPDDVDODEVAEYFVRDRKRLPLELTELTEEFNATVMASSIVLYFGQWAVSMA 406
 Db 334 EGPVTVVAVKANNEIVLDDTK-----DYLIEFVAPWCHGCKA 370
 QY 407 FLQSTIDVA--VKLGSTMLLTRINCADMSDCTKONVTEFPITKMKGE--NPVSY 461
 Db 371 LAPKYEELGALYAKSEFDRVIAKVD--ATANDV--PDEIQGFPIKILYPAGAKQAPVY 427
 QY 462 AGMLGTXDLKFTQNLNRSYVNTSIOEAE 493
 Db 428 SGRTEVDLIKFIENG-KYKAIS--EDAE 456

RESULT 4

AAH73963 ID AAR73963 standard; Protein; 505 AA.

XX AAR73963;

DT 12-JAN-1996 (first entry)

XX Humicola insolens protein disulphide isomerase.

KW Humicola insolens; protein disulphide isomerase; Bacillus brevis;
 promoter; recombinant production; fungal protein.

XX Humicola insolens.

XX Key Location/Qualifiers

XX Peptide 1..20

XX Peptide 21..505

XX Peptide /label-mat_peptide

XX JP07107980-A.

XX 25-APR-1995.

XX 30-SEP-1993; 93JP-0245279.
 PF 30-SEP-1993; 93JP-0245279.
 PR (TOYOTA CHUO KENKYUSHO KK.
 XX (UTAKA/) UTAKA J.
 PA WPI; 1995-190180/25.
 DR N-PSDB; AA087670.
 XX
 PT Fungal protein disulphide isomerase encoded by recombinant DNA
 PT contg. Bacillus brevis promoter - useful for efficient production of
 PT large quantities of the enzyme.
 PS Claim 2; Pages 8-11; 13pp; Japanese.
 CC AA087670 encodes AAR3963 Humicola insolens protein disulphide
 CC isomerase (PDI). By linking a Bacillus brevis promoter to the cDNA,
 CC the fungal PDI can be expressed in B. brevis cells. The
 CC recombinant PDI can then be isolated from the culture medium, with
 CC a yield as high as 500mg/l of culture supernatant.
 CC
 XX Sequence 505 AA:

Query Match 4.5%; Score 185; DB 16; Length 505;
 Best local similarity 21.9%; Pred. No. 2e-07;
 Matches 112; Conservative 79; Mismatches 213; Indels 108; Gaps 22;

OY 15 VIMCIFYMPTVNSLPESLPKYSTLOPGLLENEAVRPLODYGISAVAKVCKEISRY 74
 DB 40 LVLAEEFAPWCGCHKALAP-----EYEEAATTLKEKNIKIAKVDT----- 80
 OY 75 CGKEKDLMAVLEFKGNILLREF-----PTDTLFDVNAIVAHVLPALL--FSEVYITN 125
 DB 81 --EETDLCOOHGVEGYPTLKVFRLDNVSPYKGRKAATSYMIKOSLPVASEVTK-DN 137
 OY 126 LEDLONIENALKKANITISYVAIGIPEHRAVMEAGFYGTYYOVLTEITALLSIGS 185
 DB 138 LEEFKKADKA-----VLVAAYDASDKASSEVFTOVAEKLRDNYPRGSSDAAALAEAGV 191
 OY 186 EDEYVAHLYEFHCKLVLDLTOQCRRTIMEOPRTLTNIHLFIKTMKAPLITEVAEDPOVS 245
 DB 192 KAP-----AIVLYKDFDEGKAVFSEKFEVEAIEKPAKGARPLIGEIG--PEYVS 239
 OY 246 TVHLQGLPLVLFVISOA-----TYEADRTAEW--VAMRLIG-KAGVILLR 290
 DB 240 D-YMSAGIPLAIFATTAERKELSDKLPRIAEAGVINFGTIDAKAGAHGNNLTKT 298
 OY 291 DSLLEV---NIPODANVVERKRAEGVPVEFLVHDVLLIISHVENNMHIEIQEDEDNDM 346
 DB 299 DKFPAPAIQEVANOKFPPEQERE-ITFEAIKAFVDDFVAGKIEPSIKSEPIPEKO---- 353
 OY 347 EGPDIIVODEVAETVFRDRKRLPLELVELTEEFNATVMSDSIVLFYAGQVAVSMA 406
 DB 354 EGPVTVVAKNNEIYLDTK-----DVLIEFYAPWCGCHCKA 390
 OY 407 FLQSYIDVA---VKLGSTMLLTRINCADWSDVCTKONVTEPIIKMYKGE--NPVS 461
 DB 391 LAKYIELGALYAKSEFKBRVYIAKV-DATANDV--PDEIQGPTTIKLIPAGAKGQPVY 447
 OY 462 AGMLGTRDKLLEFIQNRISYPNVITSIQEAE 493
 DB 448 SGSRVEEDLIKFIENG-KYKAIS--EDAE 476

RESULT 5
 AAR63807
 ID AAR63807 standard; Protein; 505 AA.
 XX
 AC AAR63807;
 XX
 DT 29-JUN-1995 (first entry)

XX DE Protein disulphide isomerase (PDI).
 XX
 XX Protein disulphide isomerase; PDI; Humicola insolens;
 XX
 XX sulphhydryl group containing reducer.
 XX
 XX Humicola insolens.
 XX
 XX
 XX Key Location/Qualifiers
 XX Peptide 1..20
 XX FT /label= sig_peptide
 XX
 XX JP06253857-A.
 XX
 XX
 XX PD 13-SEP-1994.
 XX
 XX PF 04-MAR-1993; 93JP-0044014.
 XX
 XX PR 04-MAR-1993; 93JP-0044014.
 XX
 XX PA (TOYOTA CHUO KENKYUSHO KK.
 XX
 XX WPI; 1994-328999/41.
 XX
 XX DR N-PSDB; AA074310.
 XX
 XX
 XX PT Protein disulphide isomerase (PDI) gene - encodes a PDI showing
 XX increased stability, and tolerance against a sulphhydryl gp-contg.
 XX reducer
 XX
 XX PS Claim 1; Page 12; 15pp; Japanese.

XX
 CC AA073410 encodes AAR63807 Humicola insolens protein disulphide
 CC isomerase (PDI), which was used in the construction of a vector.
 CC The vector was used to transform suitable host cells enabling
 CC them to produce PDI. PDI produced in this way had increased
 CC stability and tolerance to sulphhydryl group containing reducers,
 CC compared to PDI obtained from mammals and yeast.

XX Sequence 505 AA:
 Query Match 4.3%; Score 180; DB 15; Length 505;
 Best local similarity 21.7%; Pred. No. 5.7e-07;
 Matches 111; Conservative 80; Mismatches 213; Indels 108; Gaps 22;

OY 15 VIMCIFYMPTVNSLPESLPKYSTLOPGLLENEAVRPLODYGISAVAKVCKEISRY 74
 DB 40 LVLAEEFAPWCGCHKALAP-----EYEEAATTLKEKNIKIAKVDT----- 80
 OY 75 CGKEKDLMAVLEFKGNILLREF-----PTDTLFDVNAIVAHVLPALL--FSEVYITN 125
 DB 81 --EETDLCOOHGVEGYPTLKVFRLDNVSPYKGRKAATSYMIKOSLPVASEVTK-DN 137
 OY 126 LEDLONIENALKKANITISYVAIGIPEHRAVMEAGFYGTYYOVLTEITALLSIGS 185
 DB 138 LEEFKKADKA-----VLVAAYDASDKASSEVFTOVAEKLRDNYPRGSSDAAALAEAGV 191
 OY 186 EDEYVAHLYEFHCKLVLDLTOQCRRTIMEOPRTLTNIHLFIKTMKAPLITEVAEDPOVS 245
 DB 192 KAP-----AIVLYKDFDEGKAVFSEKFEVEAIEKPAKGARPLIGEIG--PEYVS 239
 OY 246 TVHLQGLPLVLFVISOA-----TYEADRTAEW--VAMRLIG-KAGVILLR 290
 DB 240 D-YMSAGIPLAIFATTAERKELSDKLPRIAEAGVINFGTIDAKAGAHGNNLTKT 298
 OY 291 DSLLEV---NIPODANVVERKRAEGVPVEFLVHDVLLIISHVENNMHIEIQEDEDNDM 346
 DB 299 DKFPAPAIQEVANOKFPPEQERE-ITFEAIKAFVDDFVAGKIEPSIKSEPIPEKO---- 353
 OY 347 EGPDIIVODEVAETVFRDRKRLPLELVELTEEFNATVMSDSIVLFYAGQVAVSMA 406
 DB 354 EGPVTVVAKNNEIYLDTK-----DVLIEFYAPWCGCHCKA 390
 OY 407 FLQSYIDVA---VKLGSTMLLTRINCADWSDVCTKONVTEPIIKMYKGE--NPVS 461

Db 391 LAPYEELGALYAKSEKDRVIAKVD-ATANDV--PDEIOGFPIIKLYPAGAKGQPTY 447
QY 462 AGMLGTDLKFIOLNRIISYPVNTSIOAE 493
Db 448 GSRVTEDLIKFIAENG-KYKAIS--EDAE 476

RESULT 6
ABB62855
ID ABB62855 standard; Protein; 496 AA.
AC ABB62855;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 15357.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL06958.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 15357; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 496 AA;
SQ

Query Match 4.3%; Score 177; DB 22; Length 496;
Best Local Similarity 19.9%; Pred. No. 1e-06;
Matches 99; Conservative 96; Mismatches 180; Indels 122; Gaps 25;

QY 373 ELTVELTEETNATV-----MASDSIVL--FYAGQAVSMFLQSYIDVAVKL-KGST 423
Db 20 EAEVVEEGVULVATDNFKQLADNEFVLVEFYAPWCGHCKALAEYAKAQQLEKESP 79
QY 424 MLTRINCADSDVCTKONTVEFPIIKMYKKGENVSYAGMLGTDLKLFIOLNRIISYPV 483
Db 80 IKLAVADVTEGELAEQAVRGPPLTKFRRSG-SPVEYSGGQADIIAMVTKTKGPPAK 138
QY 484 NITSIOAEVYISGELVYDLIIYSVSVLGFSPTMKTAKEPSENGNYLKQYVITGIYS 543

Db 139 DLTSVADAEQFL-----KD-----NEIAIGFFPKDLESEBAKTTFTKVNALDSFE-GVSS 188
QY 544 EEDVLLSTTKRYAASLPALLIARHTEGKTESIPLASTHADIYQIITDALLEMPEITVEN 603
Db 189 NADVI---AKTEAKDNGVLLKRFDDK-----KSYFE--GELNEEN 224
QY 604 LPSYFRLQK-PLLIIFSDGTVPQY---KKAILTLYKOKYLDSETPCWLNLKNTPVGNG 658
Db 225 LKFAQVQSLPLIVDFNHESASKIFGSIKSHLFFVSE-----GG 266
QY 659 ILRAYEDLPPLPL-----LVLVNLHSGQ---VPAF----- 687
Db 267 HIEKYVDPLKEIAKKYRDDILFVTISDEDEHTRIFEEFGMKKEVPYTRILIKLEDMAK 326
QY 688 --PSDQATIEENLVMLKK-LEAGLENHTT--LPAQEN-KRP--LPAVDLSM-IDAA 737
Db 327 YKPESDDLSAETIEAFLEKFKLDGKLQHLQSELP-EDWKNPVKVLVSSNESVALDS 385
QY 738 TSOGRTRKVP---KCMKETDVOENDKEHEDKSAV-----RKPEIETLRIRKHNRSM 787
Db 386 KSVLYEFYAPMGCHCKQAPLYDQLAEKYKKNEDIVIAKMDSTANLESIKISFTPIKY 445
QY 788 FKEAEK---SFRDKEL 801
Db 446 FRKEDNKVIDENLDRTL 462

RESULT 7
AAR82801
ID AAR82801 standard; Protein; 505 AA.
AC AAR82801;
XX
XX 29-NOV-1995 (first entry)
XX
XX Bovine phospholipase C-alpha (PLC-alpha).
DE
XX
XX Phospholipase C; PLC; alpha; anti-inflammatory; cancer.
KM
XX
XX Bos taurus.
OS
XX
PN WO9508624-A.
XX
PD 30-MAR-1995.
XX
XX 22-SEP-1994; 94WO-JP01572.
PF
XX
XX 24-SEP-1993; 93JP-0238402.
PR
XX
XX (SHIO) SHIONOGI & CO LTD.
PA
XX
PI Hirai H, Hirano N;
XX
XX WPI; 1995-139593/18.
DR N-PSDB; AAT00952.
XX
XX Human phospholipase C-alpha and DNA encoding it - useful as anti
PT inflammatory
PT
XX
XX Example 1; Page 25-29; 39pp; Japanese.
PS
XX
XX AAR82801 is bovine phospholipase C-alpha (PLC-alpha). PLC-alpha
CC can be used as an anti-inflammatory and may also be used to
CC measure cancer progression.
CC
XX
XX Sequence 505 AA;
SQ

Query Match 4.2%; Score 175.5; DB 16; Length 505;
Best Local Similarity 23.3%; Pred. No. 1.4e-06;
Matches 66; Conservative 56; Mismatches 128; Indels 33; Gaps 9;

QY 376 VELTEETNATWASDS---IVLFAGQAVSMFLQSYIDVAVKLKSTMLTRINC 431

Db 28 LETLDNEESRTITDSSGLMVEFFAPWCCHKIAPYEAAATRLKG--IVPLAKVDC 85
 QY 432 ADMSDVCTKONTEPPIIKMYKKGPNVSYAGMLGTKDLKFIOLNRISYPVNITSIOEA 491
 Db 86 TANTMCKNKYGSVPTLTKIFRDGESSAVDGPRTADGIVSHLKKQAGASVPLKSEEEF 145
 QY 492 EHYLSGELXKDLILYSSSVYLGFSPITKTAKEDFSEAGNYLKGVIYTTIYSEEDVLLS 551
 Db 146 EKFIIDK-----DASVGFPPKDLFSEAHSEFLKASNLRD--NYRFAHTNYESLV 193
 QY 552 TKYASLPALLAR--HTEGKIESIPLASTHAQ---DIVQITTDALLEMPEITVENLP 605
 Db 194 NKYDDDGEGITLFRPSHLTKFEDKTVATEQKMTSGKRIQENITFICHTMEDND 253
 QY 606 SYFRLO-KPLLIFSD-----GTVPNPKYKAILTLVKOKYLD 642
 Db 254 --LLOGKDLIAYYDVDEKNAKSNYWRNRVMVAKKFLDA 293

RESULT 8

AAR69506
 ID AAR69506 standard; Protein: 511 AA.

AC AAR69506;

DT 18-AUG-1995 (first entry)

DE Aspergillus sp. recombinant protein-disulfide-isomerase.

KW Protein-disulfide-isomerase; enzyme; scleroprotein;
 KM filamentous fungus.

XX Aspergillus.

XX W09500636-A.

XX PD 05-JAN-1995.

XX PF 28-JUN-1994; 94WO-DK00266.

XX PR 28-JUN-1993; 93DK-0000768.

XX PA (NOVO) NOVO-NORDISK AS.

XX HJort CM;

XX WPI: 1995-052073/07.

XX DR N-PSDB: AAO81934.

XX PT Protein dt-sulphide isomerase enzyme - obtd. from Aspergillus,
 PT useful for e.g. treating scleroproteins such as human hair or for
 PT fabric cleaning

XX PS Disclosure; Page 10; 61pp; English.

XX CC This sequence encodes a protein-disulfide-isomerase (PDI)

CC which is useful for treatment or degradation of scleroproteins,
 CC especially hair, skin and wool, etc. The protein is prepared

CC recombinantly from Aspergillus oryzae IFO 4177 or Aspergillus
 CC niger A524. The protein may be truncated into various peptides,
 CC all with PDI activity. This protein differs from AAR69505 in that

CC AA 511 is Ala not Glu.

XX SQ Sequence 511 AA;

Query Match 4.28; Score 172.5; DB 16; Length 511;
 Best Local Similarity 19.98; Pred. No. 2.7e-06;

Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

QY 15 VIMCFYVPYTNLSPELSPQKVFSTLQGLELNEAVPLDQYGISVAKVNCVKEE-ISR 73
 Db 48 LVLAEFAPMGCHKALAP-KY-----EQATLELKKNIPLVAVDCTEEALCR 95

QY 74 YGKEKDLKAY----LFGNILLREPPPTDLEFDNAIYAHVLFALLESEVKIT--NLE 127
 Db 96 DQGE-----GYPTLKIFRGLDAVK--PYOGAROTEAIYSVMYKSL-PAVSVPTPENLE 147
 QY 128 DLQNTENALKGANIIFFSVRAIGIPENHRAVMEAGVYGT-----TFQVLTTEIAL 180
 Db 148 EIKTMKTI-----VIGYIASDD-----QTANDITTFPESQDNTLFASTSDASTA 194
 QY 181 EISGEDEVAYHLVFFHCKLVLDLTQOCRRITLMEQPLTTLNHLFKTKAPLLEVAED 240
 Db 195 KAEVQKP-----SIVLYKDFDEKKATYDGEIEQDALLSMWKTAFTPLVGLG-- 242
 QY 241 POOVSTVHLQGLPLVFTYSQOATYADRRFAE--WVARRLKAGVLLLDLSLEVINP 298
 Db 243 PETYSQ-YITAGIPLAYIPA-ETKEEREOFTFEFKIAKHGSIINIVTI----- 290
 QY 299 QANVYFKRAEGVPVEFLVDVLDLISHVENNMHIE-----EIQDEDN----- 344
 Db 291 -----DAKLYGAHA-GNLMDSKRPAPAIQDPEKNAKYPYDQ 327
 QY 345 --DMEGPDI-----DVODEVAETVFRD--RRKRLPELVELTEEPNATVMASDSTV 394
 Db 328 SKEVAKADIGKFTLDVLDKVEPSIKSEAIPTQEGPVTVV--ASHYKDLVLNEXDV 384
 QY 395 L--FTAGQAVSMALQSTIDVAVKLKGSTMLLTRINCADMSDVCTKONTEPPIIKMY 452
 Db 385 LLEFYAPMGCHKALAPKYEELASLYKDIPEVTIAKID-ATANDV--PDSINGFPPIKLF 441
 QY 453 KKG--ENPYSAGMLGTKDLKFIOLNRISYPVNITSIOEAEEYISGE 498
 Db 442 AAGAKDSPVEYEGSRTVEDLANFVKENG-KHKVDALVEDPKRQESGD 488

RESULT 9

AAR69504
 ID AAR69504 standard; Protein: 515 AA.

AC AAR69504;

DT 18-AUG-1995 (first entry)

DE Aspergillus sp. recombinant protein-disulfide-isomerase.

KW Protein-disulfide-isomerase; enzyme; scleroprotein;
 KM filamentous fungus.

XX Aspergillus.

XX W09500636-A.

XX PD 05-JAN-1995.

XX PF 28-JUN-1994; 94WO-DK00266.

XX PR 28-JUN-1993; 93DK-0000768.

XX PA (NOVO) NOVO-NORDISK AS.

XX HJort CM;

XX WPI: 1995-052073/07.

XX DR N-PSDB: AAO81933.

XX PT Protein dt-sulphide isomerase enzyme - obtd. from Aspergillus,
 PT useful for e.g. treating scleroproteins such as human hair or for
 PT fabric cleaning

XX PS Disclosure; Page 32; 61pp; English.

XX CC This sequence encodes a protein-disulfide-isomerase (PDI)

CC which is useful for treatment or degradation of scleroproteins,
 CC especially hair, skin and wool, etc. The protein is prepared
 CC recombinantly from Aspergillus oryzae IFO 4177 or Aspergillus

CC niger A524. The protein may be truncated into various peptides,
CC all with PDI activity.

XX Sequence 515 AA;

Query Match 4.2%; Score 172.5; DB 16; Length 515;
Best Local Similarity 19.9%; Pred. No. 2.7e-06;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

```

QY 15 VIMCIFYMPTVNSLPESLPQKFFSTLPGLPELNEAVRPLODYGISVAKVNCKEE-ISR 73
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 48 LVLAFFAPWCGCHKALAP-KY-----EQAAETLKKENIPLVAVDCTEEBALCR 95
QY 74 YCGKEKDLKAY---LFGNILLREPPDTLFDVNAIYAHVLFALLSEVKYIT--NLE 127
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 96 DQGVF-----GYPTLKIRGLDAVK--PYQARQTEAIVSYWKSL-PAVSPVPENIE 147
QY 128 DLQNIENALKGANIIFSVYRAIGIPEHRAVMEAGVYGT-----TYQVLTTEIALL 180
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 148 EIKTMDKI-----VIGIASDD-----QTANDIFTFAESQORDNLFATSDASTA 194
QY 181 ESIGSEDEYVAHLVYFHCKLVLDLTOQCRRTIMEQPLTTLNHLFKTKAKALLTEVAD 240
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEDALLSWKTKASTPLVGLG-- 242
QY 241 PQGVSTVHLQGLPLVFIYSQOATYEADRRPAE--WVAWRLLGKAGVLLLRDSLEVNIP 298
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 243 PETYSG-YITAGIPLAYIPA-ETKEERQFTFEFFIAEKHKSINIYI----- 290
QY 299 QDANVYFKRAEGVPEFVLVLDVLLISHVENNMHIE-----EIQDEEDN----- 344
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 291 -----DAKLYGAHA-GNLLDPSKPPAPALQDPEKNAKYPYDQ 327
QY 345 --DMEGPDI-----DVQDEVAETVFRD--RKRKLPLELTVELTEEFNATVMA SDSIV 394
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 328 SKEVAKADIGKFIQDVLDDKVEPSIKSEAIPTQEGPVTVV--ASHYKDLVLNKKDY 384
QY 395 L--FYAGQAVSMAPLQSTIDVAVKIKGSTMLLRINCADMSDYCTKONVTEFPIIKMY 452
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 385 LLEFYAPWCGCHKALAPKYEEELASLYKDIPEVTIAKID-ATANDV--PDSITGFPTIKLF 441
QY 453 KKG--ENPVSYAGMLGTQDLKFIQNLNRSYPVNTTSIQEAEEYLSGE 498
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 442 AAGAKDSPEYEGSRTVEDLANFVKENG-KHKVDALEVDPKKEQESGD 488

```

RESULT 10

AAB49946 ID AAB49946 standard; Protein; 515 AA.

XX AAB49946;

XX 07-MAR-2001 (first entry)

XX A. oryzae protein disulfide isomerase #1.

XX Disulphide bond; allergy; protein disulfide isomerase variant; detergent;

XX food additive; cosmetic.

XX Aspergillus oryzae.

XX MO200070064-A1.

XX 23-NOV-2000.

XX 17-MAY-2000; 2000WO-DK00265.

XX 17-MAY-1999; 99DK-0000683.

XX 18-MAY-1999; 99DK-0000689.

XX 02-JUN-1999; 99US-0137068.

XX (NOVO) NOVO NORDISK AS.

PI Hjort CM;

XX WPI: 2001-070776/08.

XX N-PSDB; AAC89028.

PT Protein disulfide isomerase variant having increased reducing
PT properties and decreased redox potential than native proteins, used to
PT reduce allergenicity of allergic proteins in feed, food or cosmetic
PT products -

PS Claim 3; Page 72-74; 82pp; English.

CC The present invention provides variants of the Aspergillus oryzae protein
CC disulfide isomerase enzyme. These are capable of reducing disulphide
CC bonds in proteins, which may be the cause of allergies in humans. The
CC proteins can be used to reduce the allergenicity of foods, to treat
CC scleroproteins, in detergents, in food additives and in cosmetics.

XX Sequence 515 AA;

Query Match 4.2%; Score 172.5; DB 22; Length 515;
Best Local Similarity 19.9%; Pred. No. 2.7e-06;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

```

QY 15 VIMCIFYMPTVNSLPESLPQKFFSTLPGLPELNEAVRPLODYGISVAKVNCKEE-ISR 73
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 48 LVLAFFAPWCGCHKALAP-KY-----EQAAETLKKENIPLVAVDCTEEBALCR 95
QY 74 YCGKEKDLKAY---LFGNILLREPPDTLFDVNAIYAHVLFALLSEVKYIT--NLE 127
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 96 DQGVF-----GYPTLKIRGLDAVK--PYQARQTEAIVSYWKSL-PAVSPVPENIE 147
QY 128 DLQNIENALKGANIIFSVYRAIGIPEHRAVMEAGVYGT-----TYQVLTTEIALL 180
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 148 EIKTMDKI-----VIGIASDD-----QTANDIFTFAESQORDNLFATSDASTA 194
QY 181 ESIGSEDEYVAHLVYFHCKLVLDLTOQCRRTIMEQPLTTLNHLFKTKAKALLTEVAD 240
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEDALLSWKTKASTPLVGLG-- 242
QY 241 PQGVSTVHLQGLPLVFIYSQOATYEADRRPAE--WVAWRLLGKAGVLLLRDSLEVNIP 298
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 243 PETYSG-YITAGIPLAYIPA-ETKEERQFTFEFFIAEKHKSINIYI----- 290
QY 299 QDANVYFKRAEGVPEFVLVLDVLLISHVENNMHIE-----EIQDEEDN----- 344
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 291 -----DAKLYGAHA-GNLLDPSKPPAPALQDPEKNAKYPYDQ 327
QY 345 --DMEGPDI-----DVQDEVAETVFRD--RKRKLPLELTVELTEEFNATVMA SDSIV 394
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 328 SKEVAKADIGKFIQDVLDDKVEPSIKSEAIPTQEGPVTVV--ASHYKDLVLNKKDY 384
QY 395 L--FYAGQAVSMAPLQSTIDVAVKIKGSTMLLRINCADMSDYCTKONVTEFPIIKMY 452
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 385 LLEFYAPWCGCHKALAPKYEEELASLYKDIPEVTIAKID-ATANDV--PDSITGFPTIKLF 441
QY 453 KKG--ENPVSYAGMLGTQDLKFIQNLNRSYPVNTTSIQEAEEYLSGE 498
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 442 AAGAKDSPEYEGSRTVEDLANFVKENG-KHKVDALEVDPKKEQESGD 488

```

RESULT 11

AAE12991 ID AAE12991 standard; Protein; 515 AA.

XX AAE12991;

XX 28-JAN-2002 (first entry)

XX Human protein disulphide isomerase (PDI).

XX Heat shock protein; HSP; HSP peptide-binding fragment; HSPF; vaccine;

XX cytotoxic T cell response; hepatitis virus; herpes simplex virus;

human immunodeficiency virus; bacteria; Mycobacterias; Rickettsia; protozoa; Leishmania; Trypanosoma; intracellular parasite; Chlamydia; sarcoma; carcinoma; cancer; human; protein disulphide isomerase; PDI.

OS Homo sapiens.

XX XX

PH Key Location/Qualifiers

FT Binding-site 1..370

FT Binding-site /note= "Peptide binding domain"

FT Binding-site 204..491

FT Binding-site /note= "Peptide binding domain"

FT Binding-site 213..351

FT Binding-site /note= "Peptide binding domain"

XX XX

PN US2001034042-A1.

PD 25-OCT-2001.

XX XX

PF 12-JAN-2001; 2001US-0759010.

XX XX

PR 20-JAN-2000; 2000US-0488393.

XX XX

PA (SRIV/) SRIVASTAVA P K.

XX XX

PI Srivastava PK:

XX XX

DR WPI: 2001-656559/75.

XX XX

PT Vaccine compositions for vaccinating against cancers and infections,

PT comprises peptide-binding fragments (PBFs) of heat shock proteins

PT (HSPs) and non-covalent complexes of PBFs of HSPs and antigenic

PT molecules -

XX XX

PS Disclosure; Fig 3; 39p; English.

XX XX

CC The invention relates to pharmaceutical compositions comprising

CC peptide binding fragments of heat shock proteins (HSPs) and non-

CC covalent complexes of HSP peptide-binding fragments (HBF) in non-

CC covalent association with antigenic molecules. Vaccines comprising

CC peptide fragments of the invention may be used to stimulate an

CC immune response, in particular cytotoxic T cell responses against

CC cells infected with viruses (including hepatitis type A, B and C,

CC influenza, varicella, adenovirus, herpes simplex (HSV), type 1 and

CC type II, rinderpest rhinovirus, echovirus, rotavirus, respiratory

CC syncytial virus, mumps virus, papova virus, papilloma virus,

CC arbovirus, cytomegalovirus, echinovirus, hantavirus, coxsackie virus,

CC measles virus, rubella virus, polio virus, HIV-I, and HIV-II;

CC bacteria including (including Mycobacterias, Rickettsia, Mycoplasma,

CC Neisseria and Legionella); protozoa (including Leishmani, Kokzidia

CC and Trypanosoma) and intracellular parasites (including Chlamydia and

CC Rickettsia). The vaccines may be used to treat cancers such as human

CC sarcomas and carcinomas, pancreatic cancer, breast cancer, ovarian

CC cancer, prostate cancer, squamous cell carcinoma, basal cell carcinoma,

CC adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma,

CC papillary carcinoma, papillary adenocarcinomas, medullary carcinoma,

CC cystadenocarcinoma, bronchogenic carcinoma, renal cell carcinoma,

CC hepatoma, bile duct carcinoma, choriocarcinoma, seminoma and embryonal

CC carcinoma. The present sequence is human protein disulphide isomerase

CC (PDI), a HSP involved in folding of many disulphide-bonded proteins.

XX XX

SO Sequence 515 AA:

Query Match 4.1%; Score 169.5; DB 22; Length 515;

Best Local Similarity 20.6%; Pred. No. 5, 1e-06;

Matches 11; Conservative 89; Mismatches 193; Indels 145; Gaps 27;

OY 15 VMCIPTVNSLPFLSPQKYSTSTLOPGLEENAVRPLQDYGISVAKYCKEELSR 74

DB 44 LVLAEPFAWCGCHKALAP-KY-----EEATELKANIPLVKVDCAEE----- 87

OY 75 CGEKRLMAVLEKGNILREF-----PTDTLPVNAIVAVHALL--PSEVYITN 125

DB 88 -----DLCHSQGVEGPTLKIFRGVDSKPYQCAROTESTIVYTIKOSLPVASSVNE-EN 141

OY 126 LEDLQNIENALKKGNIIISYVAIGIPEHRAVMEAGFYGTGYFLVTEALLLESIGS 185

DB 142 LEEIKTMDKI-----VIGYIPSDQETGYQAEKKAESQRQNYLPAATDAAIAKSEGV 195

OY 186 EDVEYAHLY--FFHCKLVLD--LTQCCRTLMQPLTLINHLFITMKAPLLEVAEDP 241

DB 196 EDPSTV-LYKDFEKKAVYDGEIEQEA-----IHSWVSASTPLVGEIG--P 239

OY 242 QGVSTVHLQGLPLVFTVISOQATEYEDRRTAEW-----YAMRLLG-KAGV 285

DB 240 ETTSG-YIGAGVLAITFA-ETKEEREKTEDKPLAQKHKAQINATIDAKKFGAAGN 297

OY 286 LLLRLDSLEVINIPD-----ANVFKRAEE--GVVPEFLVLDVLLISHVENMHIEEQ 339

DB 298 LNLDSQKFPAPAFIQDPKAKNAKPYDQAKELNADVEKEFI---QDVLDGKVEPSIKSEPV 354

OY 340 EDEDNMEGPDIDVQDEVAEYFRDRKRRLPLELVELTEEFNKTVAASDSIVL--TY 397

DB 355 ESQ-----EGP-----VTV-VVAHSYKDLVIDNDKVLLEFY 385

OY 398 AGQAVSMAFLQSYIDAVAKLKGSTMLLTRINCADMSDVCFK-----QNV 443

DB 386 APWCGCHKALAPKYDELALY-----ADHPDLAAVTTAKIDATANDVPDP 432

OY 444 TEPIIKMYKKG--ENPVSYAGMLGTRDLKFTQLNRISYPVNTSIOEAEEYLSGEL 499

DB 433 TGEPTLRYPAGAKSPIDYSGSRVEDLANFVKENG-KHNVDALNVASBEQEGGDV 489

RESULT 12

AAB82539

ID AAB82539 standard; Protein; 515 AA.

XX XX

AC AAB82539;

XX XX

DT 17-SEP-2001 (first entry)

DE Human protein disulphide isomerase.

XX XX

XX Protein disulphide isomerase; PDI; human; heat shock protein; HSP;

KW immunotherapy; therapy; cancer; infection; vaccine.

XX XX

OS Homo sapiens.

XX XX

PH Key Location/Qualifiers

FT Binding-site 1..370

FT Binding-site /note= "peptide-binding domain"

FT Binding-site 5..232

FT Binding-site /note= "peptide-binding domain"

FT Binding-site 213..351

FT Binding-site /note= "peptide-binding domain"

FT Binding-site 204..491

FT Binding-site /note= "peptide-binding domain"

XX XX

PN WO200152791-A2.

XX XX

PD 26-JUL-2001.

XX XX

PF 18-JAN-2001; 2001WO-US01781.

XX XX

PR 20-JAN-2000; 2000US-0488393.

XX XX

PA (UYCO-) UNIV CONNECTICUT HEALTH CNT.

XX XX

PI Srivastava PK:

XX XX

DR WPI: 2001-457506/49.

XX XX

PT Pharmaceutical composition, used to treat or prevent infection or

PT cancer, comprises a complex comprising a heat shock protein-binding

PT fragment associated with a molecule displaying antigenicity of an

PT infectious agent or cancer cell -

XX Claim 53; Fig 3; 106pp; English.

CC The present sequence is that of human protein disulfide isomerase
CC (PDI), a heat shock protein (HSP) involved in the folding of
CC disulfide-bonded proteins. PDI contains various domains that
CC contribute to peptide binding, including a fragment that includes
CC amino acids 213-351 and is capable of binding short (10-15 amino
CC acid) peptides, and a larger fragment, including amino acids
CC 204-491, which is able to bind polypeptides of 25-50 amino acids.
CC An even larger fragment, including amino acids 1-370 is required
CC for more complex substrates, such as larger polypeptides. The
CC invention relates to complexes of peptide-binding fragments of HSPs
CC with antigenic molecules and their use in immunotherapy for the
CC treatment of infectious diseases and cancer. Claimed methods of
CC treating or preventing cancer/infectious disease involve culturing
CC a cancer cell/infected cell transformed with a nucleic acid
CC encoding a HSP peptide-binding domain, recovering complexes of the
CC HSP fragments noncovalently associated with peptides from the
CC cancer cell/infected cell, and administering the recovered
CC complexes. These methods can use PDI peptide-binding fragments
CC comprising amino acids 5-232, 213-351 and 204-491 of the present
CC sequence.

XX Sequence 515 AA;

Query Match 4.1%; Score 169.5; DB 22; Length 515;
Best Local Similarity 20.6%; Pred. No. 5.1e-06;
Matches 111; Conservative 89; Mismatches 193; Indels 145; Gaps 27;

QY 15 VMICIFWPTVNSLPESLPQYFTSLQGLELNEAVRPLDYGISVAKVCKEISRY 74
DB 44 LVLAFFAPWCGHCKALAP-KY-----BEAATLAKAKNIPVKVDTAAE----- 87
QY 75 CGKEKDLKAVLFKGNILREF-----PTDTLFDAVAIAVAVHLLPALL--FSEVKYITN 125
DB 88 -----DLCRSGVGEGYPTLKIFRGVDSKPYOGAROTESIYSYMKOSILPAVSSVNE-EN 141
QY 126 LEDLONTENALGKANTIFSVRAIGPEHRAVMEAGFVYGTQVFLVLTETALLSESGS 185
DB 142 LEEIKTMDKI-----VIGTIFSDDETQAFEFKYSQSDNRYLFAATDAAIAKSGEV 195
QY 186 EDVEYAHLY--FFHCKLVLD--LTQOCRRITMEQPLTTLNHLFKTKAKAPLLEVAEDP 241
DB 196 EGPSTV-LYKDFDEKKAAYDGEIIEQEA-----IHSWKASSTPLVGEIG--P 239
QY 242 QOVSTVHLQGLPLVFIVSQATTEADRRTAEM-----VAMRLIG-KAGV 285
DB 240 ETVSG-YIGAGVPLAIFYA-ETKEEREKYTEDEKPIAQKHKAINIATIDAKMFGAHAGN 297
QY 286 LLLRLDSLEVINIPQD---ANVVEKRAEE--GVPEFVLVDVDLIISHVENNHIEIQ 339
DB 298 LNLDSQKPPAPAIQDPKANNAYPYDQAKELNADEVEKFI---QVDLQKVEPSIKSEPPV 354
QY 340 EDEDNDMEGPDIDVQDEVAETFEFRDKRKLPLELTVLEETEFNATYMASDSIYL--FY 397
DB 355 ESQ-----EGP-----VTV--VVAHSKYDVIYNDKDYLLLEFY 385
QY 398 AGWQAVSMALQSYIDVAVKLGSTMLLTRINCADMSDVCTK-----ONV 443
DB 386 APWCGHCKALAPKYDELAAALY-----ADHPDLAAKVTIAKIDATFANDVPPPI 432
QY 444 TEFPIIKKKYG--ENPVSYAGMLCTDKLKFQIQLNRISYVNTISIOAEYSIGEL 499
DB 433 TGFPTLRLRYPGAGADSPLEYSGSRTEVEDLANFVKENG--KHNVADALNVASEFTOEGGDV 489

RESULT 13

AAB43424

ID AAB43424 standard; Protein: 544 AA.

XX AC AAB43424;
XX

DT 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO:869.

KW Human: cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
KW antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.

OS Homo sapiens.

PN WO20005350-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05882.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WP1; 2000-587533/55.

DR N-PSDB; AAC77633.

PT Novel isolated nucleic acids comprising sequences encoding peptides
useful for treating or diagnosing e.g. cancer -

PS Claim 11; Page 1424-1426; 2352pp; English.

CC AAC78467 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnery; immunomodulator;
CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

SO Sequence 544 AA;

Query Match 4.1%; Score 169.5; DB 21; Length 544;
Best Local Similarity 22.6%; Pred. No. 5.6e-06;
Matches 64; Conservative 59; Mismatches 127; Indels 33; Gaps 10;

QY 376 VELTEETFNATYMASDS-----IVLFYAGWQAVSMALQSYIDVAVKLGSTMLLTRINC 431

DB 67 LEITDDNFESRISDQSGMLMVERFAPWCGHCKRLAPLEYEAARLKG--IVLAKDKC 124

QY 432 ADMSDVCTKQVNTTEPIIKMYKKGPNPVSYAGMLCTDKLKFQIQLNRISYVNTISIOEA 491
DB 125 TANTTCKNKYGVSGVPTLKIFRDEGEAGAYDGPRTADIVS--HLKKQAGASVYPLRTEE 182

Oy		492	EELYSGELVYKDLIIYSVSVALGTSPTTKTAKKEPSEGANGLKGVITGIYSEDVILLS	551
Dd		183	E-----FKFISDKDASIVGFDDDSSEHSEFLKAASLNRD---NYRAHNVESLV	232
Oy		552	TKVAASLPALLLAR--HTEGKIETSIPIASTHAQ---DIVOITDALLEMPETIVENLR	605
Dd		233	NEYDNGEGIIIFRPSHLTNKFEDKTAVYEQMKMTSGIKKKFIQENIFGICPHMTEDNKD	292
Oy		606	SYFRHQ-KPLILFSD-----GYNAPQVKAILLLYOKIYIDS	642
Dd		293	--LIQGKDLLIAYYDYVDYEKNAGKSINYMRRNMVMAKFLDA	332
 RESULT 14 ABP41717				
ID	ABP41717	standard; Protein; 549 AA.		
AC				
XX	ABP41717:			
XX				
DT	22-AUG-2002	(first entry)		
XX				
DE	Human ovarian antigen HPDRV42, SEQ ID NO:2849.			
XX				
KW	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;			
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;			
KW	fertility; pregnancy disorder; anovulation; polycystic ovary syndrome;			
KW	PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;			
KW	Inflammatory condition; immune disorder; blood disorder;			
KW	cardiovascular disorder; respiratory disorder; neurological disorder;			
KW	gastrointestinal disorder; urinary system disorder; drug screening;			
KW	gene therapy; chromosome mapping; forensic analysis;			
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;			
KW	antiinflammatory; gynaecological; reproductive; chromosome 15q15.			
OS	Homo sapiens.			
PN	WO200200677-A1.			
XX				
PB	03-JAN-2002.			
PF	07-JUN-2001; 2001WO-US18569.			
PR	07-JUN-2000; 2000US-209467P.			
PA	(HUMA-) HUMAN GENOME SCI INC.			
PI	Blrse CE, Rosen CA;			
DR	WPI: 2002-147878/19.			
DR	N-PSDB: ABQ54794.			
XX				
PT	Isolated nucleic acid molecules encoding novel ovarian polypeptides,			
PT	useful in the prevention, treatment and diagnosis of cancer (e.g.			
PT	ovarian cancer), immune disorders, cardiovascular disorders and			
XX	neurological diseases -			
XX				
Claim 11:	SEQ ID No 2849; 2922pp; English.			
XX				
CS	The invention relates to 2175 novel human ovarian antigens (ABP41054-			
CC	ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also			
CC	comprises polypeptides 90% identical and polynucleotides 95% identical			
CC	to the sequences of the invention. The invention additionally relates to			
CC	recombinant vectors and host cells comprising human ovarian antigen			
CC	polynucleotides, antibodies against human ovarian antigens, and the use			
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,			
CC	treating, prognosing or preventing various ovary and/or breast-related			
CC	disorders. Such conditions include ovarian cancer and breast cancer, and			
CC	metastatic tumours of ovarian or breast origin, reproductive system			
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,			
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine			
CC	disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic			
CC	shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and			
CC	vaginitis), immune disorders (e.g., congenital and acquired			

Query Match	4.1%	Score 169.5	DB 23	Length 549
Best Local Similarity	22.6%	Pred. No. 5.6e-06		
Matches	64	Conservative	59	Mismatches 127; Indels 33; Gaps 10
XX	Sequence	549 AA:		
CC	immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus)			
CC	blood-related disorders (e.g., anemia), cardiovascular disorders,			
CC	respiratory disorders, neurological disorders, gastrointestinal disorders			
CC	and urinary system disorders. Ovarian antigen polypeptides and			
CC	polynucleotides may also be used in screening for compounds which			
CC	modulate ovarian antigen expression or activity. The polynucleotides may			
CC	further be used for gene therapy, chromosome mapping, in the			
CC	identification of individuals and in forensic analysis, and the			
CC	polypeptides may be used as food additives or to prepare antibodies			
CC	useful in disease diagnosis, drug targeting and phenotyping. The present			
CC	sequence represents a human ovarian antigen of the invention.			
CC	Note: The sequence data for this patent did not form part of the printed			
CC	specification, but was obtained in electronic format directly from WIPO			
CC	at ftp.wipo.int/pub/published_pct_sequences.			
XX				
QY	376 VELTEETENATWASDS---IVLFYAGQAVSMAFLQSYIDVAVKLGKSTMLLRINC 431			
DB	72 LELTDDNFESRISDPSAGLMLVEFPAPGCHCKRLAPEYEAATRLKG--IVPLAKVDC 129			
QY	432 ADMSPVCTKONTTEPIIMKYKKGENPVPYAGMLGTRKDLKFLQRLRISVPVNTISQEA 491			
DB	130 TANTWTCNKYGVSGPYTLKIFRDGEPAGAYDGPRIADGIS--HLKKQGPASVPPLRTER 187			
QY	492 EBYLSEGLYKDLILSSVAVLGFSPMTAKEDFSEAGNYLKGVYITGIYSEEDVLLS 551			
DB	188 E-----FKRISDKDASIVGFDFDSFEASSEFLKANSNLRD---NYRFATNYESLV 237			
QY	552 TKYASLPLALLAR--HREGKTESIPLASTHAQ---DIVQITDALLKEMPEITVENLP 605			
DB	238 NEYDNGEGEIIIFRPSHLTNKFEDEKTVAYTEQKMTSGKIKKFKIQENIFGICPHMEDND 297			
QY	606 SYFRLO-KPLLLFSD----GTVPDQYKKALLITLVKKYLLS 642			
DB	298 ---LIQKDLTLAYYDVDEKNAKGSNVRNRMVNAKKFLDA 337			
RESULT 15				
AAAR70329				
ID	AAAR70329 standard; Protein, 505 AA.			
XX	AAAR70329;			
AC				
XX	22-NOV-1995 (first entry)			
DT				
XX	Human phospholipase C-alpha (PLC-alpha).			
DE				
XX	Phospholipase C; PLC; alpha; anti-inflammatory; cancer.			
KW				
XX	Homo sapiens.			
SS				
XX	Key	Location/Qualifiers		
FH	Peptide	1..24		
FT	Protein	/label= signal_peptide		
FT		25..505		
FT		/label= phospholipase_C-alpha		
XX	W09508624-A.			
FN				
XX	30-MAR-1995.			
PD				
XX	22-SEP-1994; 94WO-JP01572.			
FE				
XX	24-SEP-1993; 93JP-0238402.			
PR				
XX	(SHIO) SHIONOGI & CO LTD.			
PA				
XX	Hiral H, Hlrano N;			
II				

```

XX WPI; 1995-139593/18.
DR N-PSDB; AAQ86642.
XX
XX Human phospholipase C-alpha and DNA encoding it - useful as anti
PT inflammatory
XX
XX Claim 1; Page 21-24; 39pp; Japanese.
XX
XX AAR70329 is human phospholipase C-alpha (PLC-alpha). PLC-alpha
CC is useful as an anti-inflammatory and may also be used to measure
CC cancer progression.
XX
XX Sequence 505 AA;
SQ

Query Match 4.0%; Score 166.5; DB 16; Length 505;
Best Local Similarity 22.6%; Pred. No. 9.1e-06;
Matches 64; Conservative 59; Mismatches 127; Indels 33; Gaps 10;

QY 376 VELTEEFENATVMA SDS---IVLFYAGQAVSMFLQSYIDVAVKLGSTMLLRINC 431
DB 28 LELTDDNFESRISDTGSAGMLVEFFAPWCGCHKRLAPEYEAARLKG--IVPLAKYDC 85
QY 432 ADMSDVCTKQNVTEPPIIKMYKKGENPVSYAGMLGTDLKFLQLNRTSYPVNITSIQEA 491
DB 86 TANTNTCNKYGVSGYPFLKIFRDGEAGAYDGPRTADGIVS--HLKKQAGPASYPLRTEE 143
QY 492 EEVLSGELYKDLIYSSVSVGLFSPTMKTKAKEDFSEAGNYLKGIVTGIYSEEDYLLS 551
DB 144 E-----FKRFISDKDASIVGFDDSFSEAHSEFLKAASNLRD--NYRFAHTNVESTIV 193
QY 552 TKYAASLPALLAR--HTEGKTESIPLASTHAQ---DIVOIITDALLEMPPEITVENLP 605
DB 194 NEYDNGEGITLFRPSHLTNKFEYKTVAYTEQKMTSGKIKKFIQENIFGICPHMTEDNKD 253
QY 606 SYRILQ-KPLLILFSD-----GTVPQYKKAILLVLVKQKYLDS 642
DB 254 ---LIQKDLILAYYDVYDEKDAKGSNYMRNRVMVAKKFLDA 293

```

Search completed: March 13, 2003, 17:12:58
 Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:12:17 ; Search time 18 Seconds
(Without alignments)
1319.128 Million cell updates/sec

Title: US-09-847-046-2

Perfect score: 4142

Sequence: 1 MFGFNVFRVIGSIFVIMCIF.....FKFEKSFRRDKELGCKSVN 807

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_AA:*
1: /cgn2_6/p10data/2/1aa/5A.COMB.pep:*
2: /cgn2_6/p10data/2/1aa/5B.COMB.pep:*
3: /cgn2_6/p10data/2/1aa/6A.COMB.pep:*
4: /cgn2_6/p10data/2/1aa/6B.COMB.pep:*
5: /cgn2_6/p10data/2/1aa/6C.COMB.pep:*
6: /cgn2_6/p10data/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	185	4.5	485	US-08-068-395A-1	Sequence 1, Appl
2	185	4.5	485	US-08-464-365-1	Sequence 1, Appl
3	185	4.5	505	US-08-068-395A-3	Sequence 3, Appl
4	185	4.5	505	US-08-464-365-3	Sequence 3, Appl
5	175.5	4.2	505	US-08-627-907A-4	Sequence 4, Appl
6	172.5	4.2	511	US-08-557-122A-4	Sequence 4, Appl
7	172.5	4.2	511	US-08-262-666-4	Sequence 4, Appl
8	172.5	4.2	515	US-08-557-122A-3	Sequence 3, Appl
9	172.5	4.2	515	US-08-557-122A-34	Sequence 3, Appl
10	172.5	4.2	515	US-09-262-666-3	Sequence 3, Appl
11	172.5	4.2	515	US-09-262-666-34	Sequence 3, Appl
12	166.5	4.0	505	US-08-627-907A-2	Sequence 2, Appl
13	152.5	3.7	530	US-08-557-122A-35	Sequence 35, Appl
14	152.5	3.7	530	US-09-262-666-35	Sequence 35, Appl
15	150	3.6	521	US-09-262-666-32	Sequence 32, Appl
16	150	3.6	521	US-09-262-666-32	Sequence 32, Appl
17	138	3.3	522	US-09-368-588-2	Sequence 2, Appl
18	137	3.3	509	US-08-557-122A-29	Sequence 29, Appl
19	137	3.3	509	US-08-262-666-29	Sequence 29, Appl
20	136	3.3	466	US-08-984-919A-33	Sequence 33, Appl
21	136	3.3	470	US-08-984-919A-55	Sequence 55, Appl
22	136	3.3	495	US-08-984-919A-47	Sequence 47, Appl
23	136	3.3	508	US-08-557-122A-37	Sequence 37, Appl
24	136	3.3	510	US-08-262-666-37	Sequence 37, Appl
25	136	3.3	510	US-08-441-139-20	Sequence 20, Appl
26	136	3.3	510	US-08-557-122A-30	Sequence 30, Appl
27	136	3.3	510	US-09-262-666-30	Sequence 30, Appl

28	135.5	3.3	509	US-08-557-122A-27	Sequence 27, Appl
29	135.5	3.3	509	US-09-262-666-27	Sequence 27, Appl
30	135	3.3	468	US-08-874-102-33	Sequence 33, Appl
31	135	3.3	472	US-08-874-102-55	Sequence 55, Appl
32	135	3.3	491	US-07-872-678-3	Sequence 3, Appl
33	135	3.3	497	US-08-874-102-47	Sequence 47, Appl
34	135	3.3	510	US-08-557-122A-28	Sequence 28, Appl
35	135	3.3	510	US-09-262-666-28	Sequence 28, Appl
36	133.5	3.2	638	US-08-557-122A-38	Sequence 38, Appl
37	133.5	3.2	638	US-09-262-666-38	Sequence 38, Appl
38	133	3.2	870	US-09-134-001C-4959	Sequence 4959, Ap
39	132	3.2	510	US-08-557-122A-36	Sequence 36, Appl
40	132	3.2	510	US-09-262-666-36	Sequence 36, Appl
41	131.5	3.2	364	US-08-650-275-4	Sequence 4, Appl
42	131.5	3.2	364	US-09-181-318-4	Sequence 4, Appl
43	130	3.1	531	US-08-923-536A-12	Sequence 12, Appl
44	129.5	3.1	281	US-08-557-122A-5	Sequence 5, Appl
45	129.5	3.1	281	US-09-262-666-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-068-395A-1
Sequence 1, Application US/08068395A
Patent No. 5496719

GENERAL INFORMATION:

APPLICANT: YAMADA, YUKIO

APPLICANT: ASAMI, OSAMU

APPLICANT: SUGIYAMA, HIDEHIKO

APPLICANT: IDEKOB, CHIE

APPLICANT: HOSHINO, RUMIHIKO

APPLICANT: HIRAI, MASANA

APPLICANT: KAJINO, TSUTOMU

APPLICANT: IMABEDA, TAKAO

APPLICANT: SARAI, KIYOKO

TITLE OF INVENTION: POLYPEPTIDE POSSESSING PROTEIN DISULFIDE

TITLE OF INVENTION: ISOMERASE ACTIVITY GENE ENCODING THE SAME AND PROCESS FOR

TITLE OF INVENTION: PRODUCING THE SAME

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/068.395A

FILING DATE: 19930527

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-135254

FILING DATE: 27-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 5-44013

FILING DATE: 04-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 5-44014

FILING DATE: 04-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5496719man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 68-228-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

```

: TELEFAX: (703) 413-2220
:
: TELEX: 248855 OPAT UR
:
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 485 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-068-395A-1

```

Query Match	4.5%	Score 185;	DB 1;	Length 485;
Best Local Similarity	21.9%	Pred No. 1.7e-08;		
Matches 112;	Conservative 79;	Mismatches 213;	Indels 108;	Gaps 22

Qy	15	VIMOFIWPYPTNLSPELSPKSEFTSLOPGLTEELNEAPRLQDYSISVAKNCEEISRY	74
Dp	20	LVAHEFAPMCGHCKALAP-----EYEAFTTLKEKNIKILAKYDC-----	60
Qy	75	CGKEKDLKATLEFGNILLREF-----PDLTLEDYNAIYAHVLFALL--FSEVKTITN	1255
Dp	61	--EETDLOQGVGVGYPTLKVRGLDNVSPKGGKAAITSYMIKOSLPAVSEVTK-DN	117
Qy	126	LEDLONTIENALKGRANIIFSYVRAIGIPENHVAWAGEFVYGTTOFQVLTETIALESIGS	1859
Dp	118	LEEKKADKA-----VLVAAYDASDKASSVETFOVAKELKDRNPFPSSSDAALAEAGV	177
Qy	186	EDVEYAHLYFHCKLVLDLTQOCRTTMEODLTTLNTHLFTKMAPLTEVAEDPOOVS	245
Dp	172	KAP-----AIYLYKDFDECKAFAFSEKEFEVEALEKEFAKTOATPLIGIG--PETYS	219
Qy	246	TVHLTOLGLPVEFIYSOOA-----TYEADRRTAEW--VAMPLLG-KAGVILLR	250
Dp	220	D-VYSAGIPLATYIFAETFAEEKEKLESDKLKPLAEQKGVINFGTIDAKAFGHAHGNLUKT	278
Qy	291	DSLEY---NIPQDANVFKRAEEGVPEVLEVLHDVLLIISHVENNMHIEIODEDNDM	346
Dp	279	DKFPAPFAIOEAVKAKKQKPEFDQEK--ITFEFAIKAFVDDFVGAKIETPSIKSEPIEKO---	333
Qy	347	EGPIDVODDEVAETVFRDRKRKLPLELJYELTEETFNATYMASDSVLVLYAGQAOVMSA	406
Dp	334	EGPITVYVAKNYNETIVLDTK-----DVLEIFEYAPMGCHCKA	3707
Qy	407	FLQSYIDVA--VKLGSTWLLTRINCADMSDYCTKQNTYEPILIMYKRG--NPNYS	461
Dp	371	LAPYEEIGALYASSEKDRVVIKVD-ATANDV--DDEIOGFPTIKLYPAGAKGQGVTY	427
Qy	462	AGMGTGDKLLKFIQIOLNRIISYPVNITSIOQAAE	493
Dp	428	SGSRTVEDLKFEIAENG-KTKAALS--EDDEE	456

RESULT 2
 US-08-464-365-1
 : Sequence 1, Application US/08464365
 : Patent No. 5700659
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: YAMADA, YUKIO
 : APPLICANT: ASAMI, OSAMU
 : APPLICANT: SUGIYAMA, HIDEHIKO
 : APPLICANT: IDEKOKA, CHIE
 : APPLICANT: HOSHINO, FUMHIKO
 : APPLICANT: HIRAI, MASANA
 : APPLICANT: KAJINO, TSUTOMU
 : APPLICANT: IMAEDA, TAKAO
 : APPLICANT: SARAI, KIYOKO
 :
 : TITLE OF INVENTION: POLYPEPTIDE POSSESSING PROTEIN DISULFIDE
 :
 : TITLE OF INVENTION: ISOMERASE ACTIVITY GENE ENCODING THE SAME AND PROCESS FOR
 :
 : TITLE OF INVENTION: PRODUCING THE SAME
 :
 : NUMBER OF SEQUENCES: 10
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 :
 : ADDRESSEE: P.C.
 :
 : STREET: 1755 S. Jefferson Davis Highway, Suite 400

```

1      CITY: Arlington
2      STATE: Virginia
3      COUNTRY: U.S.A.
4      ZIP: 22202
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: Floppy disk
8      COMPUTER: IBM PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: Patentln Release #1.0, Version #1.25
11
12     CURRENT APPLICATION DATA:
13     APPLICATION NUMBER: US/08/464,365
14     FILING DATE:
15     CLASSIFICATION: 435
16
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: JP 4-135254
19     FILING DATE: 27-MAY-1992
20
21     PRIOR APPLICATION DATA:
22     APPLICATION NUMBER: JP 5-44013
23     FILING DATE: 04-MAR-1993
24
25     PRIOR APPLICATION DATA:
26     APPLICATION NUMBER: JP 5-44014
27     FILING DATE: 04-MAR-1993
28
29     ATTORNEY/AGENT INFORMATION:
30     NAME: Oblon, No. 5700659man F.
31     REGISTRATION NUMBER: 24,618
32     REFERENCE/DOCKET NUMBER: 68-228-0
33
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE: (703) 413-3000
36     TELEFAX: (703) 413-2220
37     TELE: 248855 OPAT UR
38
39     INFORMATION FOR SEQ ID NO: 1:
40     SEQUENCE CHARACTERISTICS:
41     LENGTH: 485 amino acids
42     TYPE: amino acid
43     TOPOLOGY: linear
44
45     MOLECULE TYPE: protein
46
47     US-08-464-365-1

```

Query Match	4.58;	Score 185;	DB 1;	Length 485;
Best Local Similarity	21.9%;	Pred. No. 1.7e-08;		
Matches 112;	Conservative 79;	Mismatches 213;	Indels 108;	Gaps 22;

```

Oy 15 VIMCIEWPTJNSLPELSPKUYFTSTJORGLELMEVNRPLDQVCSIAKVCNKEEISRY 74
Db 20 LVLEDFPAPWCCHCKALAP-----EYEAATTLKEKNIKIAKDCT-----60
Oy 75 CGKEKDLKAYLFKGNILLREF-----PTDLEFDNAIVANHLFALL--FSEVUYITN 125
Db 61 --EETDLOQGVGEVGYPTLKVFRGLDANVSPYKGQRKAALITSMYKQSLPASEVTK-DN 117
Oy 126 LEDQNTENALKGANITFESVYRAGIRPENHAAVGEVGVGTYQVPLTEITLALBESIGS 185
Db 118 LEEERKKADKA-----LVAVYVDSDKASSEVTPQVAKELSDNPFESSSDAALAEAGV 171
Oy 186 EDVEYANLFEFNCKLVLDLTOQCRRTTMEQPLTLNHLFKTKAKDALTEVADPQVDS 245
Db 172 KAP-----ALVLYKPDDESKAVFSEKFEVEALEKFAKTGATPLDIGIG--PETYS 219
Oy 246 TVHLQTLGFLVFIYSQA-----TYEDRRTAEV--VANKLLG-KAGVLLLR 290
Db 220 D-VYNSAGIPLAIVIAETAEEERKELSDKLPRAEQRQVINFGTIDAAFAGNHNLKLT 278
Oy 291 DSLFV-----NIPQANVYFKRAEGVFEVFLVLDVLLISHVENNNHIEIOEDENDM 346
Db 279 DKFAPALQEVAKKQKRPFODEK--TFEALKAIVDDFVAGKIEPISKSERIPKQ-----333
Oy 347 EGPDIIDVODEVAETFEVRDRKKRLLELTVELTEETENATVMSDSITVLEFYAGQAVSMA 406
Db 334 EGPITVVAAKNYNEITVLDTK-----DVLTEFYAPWCCHCKA 370
Oy 407 FLQSYIDVA---VLTQKSTIMLLRLRICADMSDVCSTQNTEPPIIMKYKGE--NNVSY 461
Db 371 LAPKTEEGALYALSEKDRVIAKV-DATAND--PDEIOGFETIKLTVIYAGAKGQRPVY 427

```


APPLICATION NUMBER: JP 5-44013
FILING DATE: 04-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-44014
FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBILO, No. 5700659man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 68-228-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-464-365-3

Query Match 4.5%; Score 185; DB 1; Length 505;
Best Local Similarity 21.9%; Pred. No. 1,9e-08;
Matches 112; Conservative 79; Mismatches 213; Indels 108; Gaps 22;

15 VMCIPTVMTVNSLPKSPKFFSTLQPGLEINEAVRPLODYGISVAVNCVEISRY 74
40 LVLAEPFAWCGCHKALAP-----EYEEAATTLKERNIKLAKVDT----- 80
75 CGKENDMAYLEFGNILLREF-----PTDTLFVNNAIVAHVLEALL--FSEVYITN 125
81 --EETDLCOOHGEGYPTLKVFGRGLDNVSPKQORRAAITSYMIQSIPAYSEVRK-DN 137
126 LEDLQNIENALGKANIIFSVRAIGIPERRAVMEAGFYGTGYOPVLTEILLESIGS 185
138 LEEFKKADK-----VLVAAYDASDKASSEVFTQYAEKLRNDYPPRGSSDALAAEGV 191
186 EDVEYAHLYFPHCKLVLDLTQOCRTLMROPITTLNHLFIKTKAPLITEVAEDPOQVS 245
192 KAP-----AIVLYKDFDEGKAVFSEKEFEVEAIEKFAKTGATPLIGEIG--PEYVS 239
246 TVHILQGLPLVIFSOA-----TYEADRTNEM--VAMRLIG-KAGVLLLR 290
240 D-YMSGIFLATIIFATAERKELSKLPILAEQGVINFGITDKAGAHAGNINLMT 298
291 DSLELV---NIPDANVVEKRAEEGVPELVLDYDLIISHVENMHIETIOEDENDM 346
299 DKPFAIQEVAKNQKFPDQKE-ITFEAKAFVDDFVAGKIEPSIKSEPIPEKQ----- 353
347 EEPDIDVODEVAEYFRRKRKLPLELTVELTEETFNATWASDSIVLFYAGWQAVSMA 406
354 EGPVTVVVAKNYNEIYLDOTK-----DVLIEFYAWCGCHCKA 390
407 FLOSYIDVA---VKLGSTMLLTRINCADMSDYCTKQNVTEPILIKMKKG- NPVS 461
391 LAPKEELGALYAKSEFKRVVIAYKD--PDEIQGPITIKLYPAGAKGQPVY 447
462 AGMLGTDKLKFQTLNRISSPVNITSIOEAE 493
448 SGRSTVEDIKRTIENG-KYKAIS--EDAE 476

RESULT 5
US-08-627-907A-4
Sequence 4, Application US/08627907A
Patent No. 6060302

GENERAL INFORMATION:
APPLICANT: HIRANO, Naoto
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE C-ALPHA AND DNA
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE C-ALPHA AND DNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: C/O FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/627,907A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-238402
FILING DATE: 24-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP94/01572
FILING DATE: 22-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: HALEY Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-627-907A-4

Query Match 4.2%; Score 175.5; DB 3; Length 505;
Best Local Similarity 23.3%; Pred. No. 1.4e-07;
Matches 66; Conservative 56; Mismatches 128; Indels 33; Gaps 9;

376 VELTEETFNATWASDS---ILFYAGWQAVSMAPLOSVIDVAVNKLKSTMLTRINC 431
28 LELTDNFSRIITDISSGLMVEFPAPCGCHKLAPYEAAATRLKG--IYPLAKVVC 85
432 ADMSDVCTKQNVTEPFIIMYKGENPVSYAGMLGTDKLKFQTLNRISSPVNITSIOEA 491
86 TANTNTCNKRYGVSYPYTLKIFRQGESGAYDGPRTADGIVSHUKQAGPASYPLKSEEF 145
492 EETLSELKDLIYSSVYGLFSPMTAKEDPSEAGNYLKGYITIGIYSEEDVLLLS 551
146 EKFIISK-----DASVGFEXKDFSEAHSEFLKASNLND--NYRFAHTNYESLV 193
552 TRYASLPAALLAR--HTGKTESIPLASTHAO---DIVQITDALLMPFEIYENLP 605
194 NKYDDGEGITLFRSHLTNKFEDKVAITTEOKMISGKIKRIGQENIFGICPHMTEDNDK 253
606 SYFRLO-KPILLIFSD-----GTVPQYKKAILLTVKQYLLS 642
254 ---LLOGRDLLATAYVDYDEKNAKGSNRYNRVMYAKKFLDA 293

RESULT 6
US-08-557-122A-4
Sequence 4, Application US/08557122A
Patent No. 5879664

GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5879664 No. 5879664disk of No. 5879664th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York

```

; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-557-122A-4

```

Query Match 4.2%; Score 172.5; DB 2; Length 511;

Best Local Similarity 19.9%; Pred. No. 2.7e-07; Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

```

QY 15 VIMCIFYMPTVNSLPESLPQKFFSTLPGLLEAVAPLDDYGSIVAKVCKEE-ISR 73
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 48 LVLAFFAPWCGHCKALAP-KY-----EQAAELKEKNIPLVKVCTEEEALCR 95
QY 74 YCGEKDLMKAY----LFKGNILIREPPTDLPVNAIVAHVLAFLFSEKYYIT--NLE 127
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 96 DQGEV-----GYPLKIFRGDAVK--PYGARGQTEAIVSVWKOST--PAVSPTPENLE 147
QY 128 DLQIENNAKGNIIISYVAIGIPEHRAVMEAGFYVGT-----TYQVLTTEIALL 180
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 148 EIKTMDKI-----VVIQYIASDD-----QTANDITTTAESORDNYLFAAATSDASIA 194
QY 181 ESIGSEDEVAYHLYFFHCKLVLDLTOCCRTLMEOPLTTNINHLFIKTMKAPLLEVAED 240
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 KAEVVKRP-----SYLVKDFDEKKAITYGGEIQDALLSKWTASTPVLGELG-- 242
QY 241 PQOVSTVHLQGLPLVFIVSQOATYEADRRPAE--WVAMRLGKAGVLLLRDSLLENVIR 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 PETYSG-YITAGIPLAYIFA-ETKEERQGFEEFKFIKHKGSINIYTI----- 290
QY 299 QDANVVKRAEGVPEFLVHDVLLISHVENNMHIE-----EIQEDSDN----- 344
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 291 -----DAKLYGAHA-GNINLDPSPFPAPAIODPPKNAKYPYDQ 327
QY 345 --DMEGPD-----DVODDEVAETVFRD--RKRRLPELTVETLEETFNATVNASDSIV 394
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 SKEVKADIGKFIQDVLDKVEPSIKSEAIPEIQEGPTVVV--AHSYKDLVDNEKD 384
QY 395 L-FYAGQAVSMAFLQSYIDVAVKLGKSTMLLTRINCADMSVCTKQNTTEPPIIKMY 452
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 385 LLEFYAPWCGHCKALAPKYEELASLYKDIPVTLAKID-ATANDV--PDSITGPTIKLF 441
QY 453 KKG--ENPVSTAGMLGKTKDLKFIQLNRISYPAVITSIQEAEFLSGE 498
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 442 AAGAKDSPVEYEGSRTEVDLANFVKENG-KHKVDALVDPKKEOESGD 488

```

RESULT 7
US-09-262-666-4
; Sequence 4, Application US/09262666
; Patent No. 6346244
; GENERAL INFORMATION:

```

; APPLICANT: Hjort, Carsten Mølland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-262-666-4

```

Query Match 4.2%; Score 172.5; DB 4; Length 511;

Best Local Similarity 19.9%; Pred. No. 2.7e-07; Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

```

QY 15 VIMCIFYMPTVNSLPESLPQKFFSTLPGLLEAVAPLDDYGSIVAKVCKEE-ISR 73
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 48 LVLAFFAPWCGHCKALAP-KY-----EQAAELKEKNIPLVKVCTEEEALCR 95
QY 74 YCGEKDLMKAY----LFKGNILIREPPTDLPVNAIVAHVLAFLFSEKYYIT--NLE 127
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 96 DQGEV-----GYPLKIFRGDAVK--PYGARGQTEAIVSVWKOST--PAVSPTPENLE 147
QY 128 DLQIENNAKGNIIISYVAIGIPEHRAVMEAGFYVGT-----TYQVLTTEIALL 180
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 148 EIKTMDKI-----VVIQYIASDD-----QTANDITTTAESORDNYLFAAATSDASIA 194
QY 181 ESIGSEDEVAYHLYFFHCKLVLDLTOCCRTLMEOPLTTNINHLFIKTMKAPLLEVAED 240
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 KAEVVKRP-----SYLVKDFDEKKAITYGGEIQDALLSKWTASTPVLGELG-- 242
QY 241 PQOVSTVHLQGLPLVFIVSQOATYEADRRPAE--WVAMRLGKAGVLLLRDSLLENVIR 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 PETYSG-YITAGIPLAYIFA-ETKEERQGFEEFKFIKHKGSINIYTI----- 290
QY 299 QDANVVKRAEGVPEFLVHDVLLISHVENNMHIE-----EIQEDSDN----- 344
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 291 -----DAKLYGAHA-GNINLDPSPFPAPAIODPPKNAKYPYDQ 327
QY 345 --DMEGPD-----DVODDEVAETVFRD--RKRRLPELTVETLEETFNATVNASDSIV 394
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 SKEVKADIGKFIQDVLDKVEPSIKSEAIPEIQEGPTVVV--AHSYKDLVDNEKD 384
QY 395 L-FYAGQAVSMAFLQSYIDVAVKLGKSTMLLTRINCADMSVCTKQNTTEPPIIKMY 452
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 385 LLEFYAPWCGHCKALAPKYEELASLYKDIPVTLAKID-ATANDV--PDSITGPTIKLF 441

```

QY 453 KKG-ENPVSYAGMLGTCKDLKFIQNLNRSYPVNITSIOAEELYSGE 498
DB 442 AAGAKDSPVEYEGSRTEVDLANFVKENG-KHKVDALEVPKKEQESOD 488

RESULT 8

US-08-557-122A-3
Sequence 3, Application US/08557122A
Patent No. 5879664
GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 515 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-122A-3

Query Match 4.2%; Score 172.5; DB 2; Length 515;
Best Local Similarity 19.9%; Pred. No. 2,8e-07;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

QY 15 VIMCIFYMPTVNSLPESLQKYFSTLQPGLEELNEAVRPLQDYGISVAKVNCVKEE-ISR 73
DB 48 LVLAEEFAWCGCHKALAP-KY-----EQATLKKENINPLVAVDCTEEBALCR 95
QY 74 YGSEKEDLMKAY----LFGNLLREFPDTLFDVNAIYAHVLFALLSEVKYIT--NLE 127
DB 96 DQGE-----GYPLKIFFGDAVK--PYOGARQTEAIVSYWKSL-PAVSPVTPEENLE 147
QY 128 DLQNIENALKGANIIFFSVRAIGIPENRAVMEAGFVYGT-----TYQFVLTEIAL 180
DB 148 EIKTMDKI-----VIGIYASDD-----QTANDIFTFAESQRDNYLFAATSDASIA 194
QY 181 ESIGSEDEVYAHLYEFHCKLVLDLTQOCRRITLMEQPLTTLNHLFIKTKAKAPLITEVAED 240
DB 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEQDALLSWKTAFTPLVGLG-- 242
QY 241 POOVSTVHLQGLPLVEIYSQATYEAADRTRAE--VWAMRLKAGAVLLLRDSLEVNIP 298
DB 243 PETYSQ-YTTAGIPLAYITA-EIKEREQPTFEFKIAEKHGSINIVIT----- 290
QY 299 QDANVYFKAEBGVPEVFLVDLIDVDLIISHVENNHIE-----EIQDEEDN----- 344
DB 291 -----DAKLYGAHA-GNLLDPSKPPAFRIQDPEKNAKYPYQD 327
QY 345 --DMGSPDI-----DVQDEVAETVFRD--RRKRLPLELTVLELTETFNATVMASSDITV 394

DB 328 SKEVKAADIGKFIQVDLDDKVEPSIKSEAIPEQOEGPVTVVV--ASHYKDLVLDNEKDV 384
QY 395 L--FYAGQVAVSAFLQSYIDVAVKLGKSTMLTRINCADMSDYCTKQNVTEFFITKAY 452
DB 385 LLEFYAPWCGCHKALAPKYEELASLYKDIPETVIKID-ATANDV--PDSITGFPIIKLF 441
QY 453 KKG-ENPVSYAGMLGTCKDLKFIQNLNRSYPVNITSIOAEELYSGE 498
DB 442 AAGAKDSPVEYEGSRTEVDLANFVKENG-KHKVDALEVPKKEQESOD 488

RESULT 9

US-08-557-122A-34
Sequence 34, Application US/08557122A
Patent No. 5879664
GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 515 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-557-122A-34

Query Match 4.2%; Score 172.5; DB 2; Length 515;
Best Local Similarity 19.9%; Pred. No. 2,8e-07;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

QY 15 VIMCIFYMPTVNSLPESLQKYFSTLQPGLEELNEAVRPLQDYGISVAKVNCVKEE-ISR 73
DB 48 LVLAEEFAWCGCHKALAP-KY-----EQATLKKENINPLVAVDCTEEBALCR 95
QY 74 YGSEKEDLMKAY----LFGNLLREFPDTLFDVNAIYAHVLFALLSEVKYIT--NLE 127
DB 96 DQGE-----GYPLKIFFGDAVK--PYOGARQTEAIVSYWKSL-PAVSPVTPEENLE 147
QY 128 DLQNIENALKGANIIFFSVRAIGIPENRAVMEAGFVYGT-----TYQFVLTEIAL 180
DB 148 EIKTMDKI-----VIGIYASDD-----QTANDIFTFAESQRDNYLFAATSDASIA 194
QY 181 ESIGSEDEVYAHLYEFHCKLVLDLTQOCRRITLMEQPLTTLNHLFIKTKAKAPLITEVAED 240
DB 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEQDALLSWKTAFTPLVGLG-- 242
QY 241 POOVSTVHLQGLPLVEIYSQATYEAADRTRAE--VWAMRLKAGAVLLLRDSLEVNIP 298

```

Db 243 PETSG-YITAGIPLATIFA-ETKEERQFTFEFKFIKHSINITYI----- 290
Qy 299 QDANVFRAEGVVEELVLHDVLIISHVENNHIE-----EIOEDEN----- 344
Db 291 -----DAKLYGAHA-GNINLDPKFPAPAIDPEKNAPYPDQ 327
Qy 345 --DMEGPD-----DVODEVAETVFRD--RRRLPLELTVELTEFPNATVMSDSIV 394
Db 328 SKEVKAKDIGKFIQDVLDDKVEPSIKSEAIPETOGPTVVV--ASHYKDLVLNEDV 384
Qy 395 L-FYAGQAVSAMFLQSYIDYAVKLGKSTMLLRINCADMSDYCTQNTPEPITMY 452
Db 385 LLEFYAPWCGHCKALAPRYEELASLYKDIPEVTIAKID-ATANDV--PDSITGFTIKLF 441
Qy 453 KKG--ENPVSYAGMLGTOKDLKFIQNLNRSYPVNTSIOEAEEVLSGE 498
Db 442 AAGAKDSFVEYEGSRTEVDLANFVKENG-KHKVDALVDPKKEDESGD 488

```

RESULT 10

```

US-09-262-666-3
; Sequence 3, Application US/09262666
; Patent No. 6346244
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambdiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-262-666-3

```

Query Match 4.2%; Score 172.5; DB 4; Length 515;

Best Local Similarity 19.9%; Pred. NO.2.8e-07; Mismatches 185; Indels 131; Gaps 28;

```

Qy 15 VIMCFYWPVTNLSPELSPQKFFSTLQGLLELNEAVRPLDYGISVAKVCVKE-ISR 73
Db 48 LVLAFFAPWCGHCKALAP-KY-----DQAATELKEKNIPYKVDCTEEBALCR 95
Qy 74 YCGEKDLMKAY---LFKGNILREPTDPLFDVNAVIVALLFSEVKYIT--NLE 127
Db 96 DQGE-----GYPLKIFRGIDAVK--PYOGARQTEALIVSYVXQSL-PAVSPTPEML 147

```

```

Qy 128 DLQNIENALKGKANIIFSVRAIGIPENRAVMEAGFVYGT-----TYQFVLTEALL 180
Db 148 EIKTKMDK-----VIGYIASDD-----QTANDIFTFEASORDNVLFAATSASIA 194
Qy 191 ESIGSEDEYVAHLFFHCKLVLDLITQOCHRLMEOPPLTTLNHLFIKTKKAPLLEVAED 240
Db 195 KAEGVKQF-----SIYLYKDPDEKKAYVDEIRODALLSWVKTASPTLVGELG-- 242
Qy 241 PQQVSTVLQGLPLVFIVSQATYVADRTAE--WVAMRLLGKAGVLLLLNDSLEVINP 298
Db 243 PETSG-YITAGIPLATIFA-ETKEERQFTFEFKFIKHSINITYI----- 290
Qy 299 QDANVFRAEGVVEELVLHDVLIISHVENNHIE-----EIOEDEN----- 344
Db 291 -----DAKLYGAHA-GNINLDPKFPAPAIDPEKNAPYPDQ 327
Qy 345 --DMEGPD-----DVODEVAETVFRD--RRRLPLELTVELTEFPNATVMSDSIV 394
Db 328 SKEVKAKDIGKFIQDVLDDKVEPSIKSEAIPETOGPTVVV--ASHYKDLVLNEDV 384
Qy 395 L-FYAGQAVSAMFLQSYIDYAVKLGKSTMLLRINCADMSDYCTQNTPEPITMY 452
Db 385 LLEFYAPWCGHCKALAPRYEELASLYKDIPEVTIAKID-ATANDV--PDSITGFTIKLF 441
Qy 453 KKG--ENPVSYAGMLGTOKDLKFIQNLNRSYPVNTSIOEAEEVLSGE 498
Db 442 AAGAKDSFVEYEGSRTEVDLANFVKENG-KHKVDALVDPKKEDESGD 488

```

RESULT 11

```

US-09-262-666-34
; Sequence 34, Application US/09262666
; Patent No. 6346244
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambdiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-262-666-34

```


LENGTH: 530 amino acids

TOPOLOGY: linear ;

MOLECULE TYPE: peptide

US-08-557-122A-32

Query Match	3.6%	Score 150;	DB 2;	Length 521;
Best Local Similarity	31.1%	Prod. Ws. 3.5e-05.		

Best Local Similarity	41.18	F100 (NO. 3, 08-03)
Matches	67	Conservative
	66	Mismatches
	125	Indels
	59	Gaps
	16	

Best Local Similarity	41.18	F100 (NO. 3, 08-03)
Matches	67	Conservative
	66	Mismatches
	125	Indels
	59	Gaps
	16	

QY	358	VAEIVFNRKRKKLE-LIVLEIETETNATVMASDSI-VL-FYVGQMAVNAFLQSIYDVA	41.5
Db	16	LASSVFQOQEVABEDSAVVKATDSFNEYIQSDHDLAEFPWCHCKNMAPEYKAA	75
QY	416	VKLGTSTMLTRINCADMSDVCOTKQVTEPFIIMKYKKE--NPVSACMLGTDKLRF	473
Db	76	ETLV-EKNITLAQIDCJENDOLCMEHNIPEGRPSLKITKNSDVNNSIDYEGP-RTAEAVQF	133
QY	474	IQLNRISYPAVNTSIQOAEAYLSGELYKDLLIVSSVVLGLFSEPTMTAKEDFSEAGNYL	533
Db	134	--MIKQSOFA-VAVVADPLAVLANETFEVTPYIQSGKIDADFNATF-----YSMAKKHF	184
QY	534	KGYIITGISEEDVLLSTKYAASL--PALILARHTEGKTESIPLASTHAODIVQIITDA	591
Db	185	NDYFVSAENADDDFKLSIYLPASMADEPVVYNGKKAD-----IADA	225
QY	592	LLEMFPE-IYVENPSTFRLOKPLLLFSOGTVAPQYKKAIIYL-----VKQYILDSFTP	645
Db	226	--DYFEKWLQVEALPYEGEII-----DGSVAFQYVESGLPLGYLFYNDDEELEBEYKP	274
QY	646	CWLNL--KNTPVGRGIL	660
Db	275	LFTELAKKN---RGLM	287

Db 16 LASSVFAQQEAVAPEDSAVVKLATDSFNEYIQSHDLVLAFFAPWCCHCKNMAPEYVKA 75

QY 416 VKLKSTMLTRINCADMSDVCIKONTIEPPIIKMYKGE--NPVSYAGMLGTDLLKF 473

Db 76 ETLV-EKNITLAQIDCTENQDLCKMEHNI PGFP SLKIFKNSDVNNSIDYEGP-RTAEAVQF 133

QY 474 IQLNRISYPVNITSIQEAEEYLSGELYKDLLYSSSVLGLFSPMTKAKEDFSEAGNYL 533

```

Db 134 --MIKQSPA-VAVVADLPAYLANETFTPVIVQSGKIDADFNATF-----YSMANKHF 184

```

QY 534 KGYVITGYSEEDVLLSTKYASL--PALLARHTEGKIESIPLASTHAQDIVQIITDA 591

```
Db 185 NDYDVSANADDFKLSTYLPSAMDEPVVYNGKKAD-----IADA 225
```

QY 592 LLEMFPE-ITVENLPSYFRLQKPLLLESDGTVPNQYKKAITL-----VKQYLDSETP 645

Db 226 --DVFETKQLQVEALPYFGEI-----DGSVFAQVYESGLPLGLFYNDDEELEEYKP 274

QY 646 CWLNL--KNTPVGRCIL 660

Db 275 LETTELAKKN --- RGLM 287

```
Search completed: March 13, 2003, 17:15:00
Job time : 23 secs
```

Job time : 23 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:13:27 ; Search time 16 Seconds
(Without alignments)
2324.769 Million cell updates/sec

Title: US-09-847-046-2
Perfect score: 4142
Sequence: 1 MMSGFNFRVIGISFVIMCIF.....FKFAKSGFRDKEIGSKVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCr_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	ID	Description
1	4142	100.0	807	10 US-09-847-046-2	Sequence 2, Appl
2	177	4.3	496	9 US-10-108-605-51	Sequence 51, Appl
3	169.5	4.1	515	10 US-09-759-010-8	Sequence 8, Appl
4	169.5	4.1	544	10 US-09-925-301-869	Sequence 869, App
5	132	3.2	468	10 US-09-925-300-1661	Sequence 1661, App
6	132	3.2	769	10 US-09-788-657-15	Sequence 15, Appl
7	129	3.1	396	9 US-09-854-133-436	Sequence 436, App
8	129	3.1	432	10 US-09-738-973-436	Sequence 436, App
9	128	3.1	432	9 US-09-978-295A-90	Sequence 90, Appl
10	128	3.1	432	9 US-09-978-697-90	Sequence 90, Appl
11	128	3.1	432	9 US-09-978-192A-90	Sequence 90, Appl
12	128	3.1	432	9 US-09-999-832A-90	Sequence 90, Appl
13	128	3.1	432	9 US-09-978-189-90	Sequence 90, Appl
14	128	3.1	432	9 US-10-174-590-74	Sequence 74, Appl
15	128	3.1	432	9 US-10-174-590-74	Sequence 74, Appl
16	128	3.1	432	9 US-10-176-758-74	Sequence 74, Appl
17	128	3.1	432	9 US-10-175-737-74	Sequence 74, Appl
18	128	3.1	432	9 US-10-173-706-74	Sequence 74, Appl
19	128	3.1	432	9 US-10-175-738-74	Sequence 74, Appl
				9 US-10-175-752-74	Sequence 74, Appl

20	128	3.1	432	9	US-10-176-482-74	Sequence 74, Appl
21	128	3.1	432	9	US-10-176-757-74	Sequence 74, Appl
22	128	3.1	432	9	US-10-176-913-74	Sequence 74, Appl
23	128	3.1	432	9	US-10-180-552-74	Sequence 74, Appl
24	128	3.1	432	9	US-10-180-557-74	Sequence 74, Appl
25	128	3.1	432	9	US-10-173-700-74	Sequence 74, Appl
26	128	3.1	432	9	US-10-174-572-74	Sequence 74, Appl
27	128	3.1	432	9	US-10-174-579-74	Sequence 74, Appl
28	128	3.1	432	9	US-10-176-492-74	Sequence 74, Appl
29	128	3.1	432	9	US-10-176-587-74	Sequence 74, Appl
30	128	3.1	432	9	US-10-176-588-74	Sequence 74, Appl
31	128	3.1	432	9	US-10-176-739-74	Sequence 74, Appl
32	128	3.1	432	9	US-10-176-740-74	Sequence 74, Appl
33	128	3.1	432	9	US-10-176-743-74	Sequence 74, Appl
34	128	3.1	432	9	US-10-176-488-74	Sequence 74, Appl
35	128	3.1	432	9	US-10-176-492-74	Sequence 74, Appl
36	128	3.1	432	9	US-10-176-747-74	Sequence 74, Appl
37	128	3.1	432	9	US-10-176-750-74	Sequence 74, Appl
38	128	3.1	432	9	US-10-176-985-74	Sequence 74, Appl
39	128	3.1	432	9	US-10-176-987-74	Sequence 74, Appl
40	128	3.1	432	9	US-10-176-991-74	Sequence 74, Appl
41	128	3.1	432	9	US-10-176-992-74	Sequence 74, Appl
42	128	3.1	432	9	US-10-184-658-74	Sequence 74, Appl
43	128	3.1	432	9	US-10-173-695-74	Sequence 74, Appl
44	128	3.1	432	9	US-10-173-697-74	Sequence 74, Appl
45	128	3.1	432	9	US-10-173-705-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-09-847-046-2
Sequence 2, Application US/09847046
Patent No. US20020068036A1
GENERAL INFORMATION:
APPLICANT: Hevezi, Peter
APPLICANT: Mack, David
APPLICANT: Gish, Kurt C.
APPLICANT: Wilson, Keith E.
TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF PROSTATE AND/OR BREAST CANCER,
TITLE OF INVENTION: COMPOSITIONS, AND METHODS OF SCREENING FOR PROSTATE CANCER, AN
FILE OF INVENTION: MODULATORS
FILE REFERENCE: A-69199-1/DJB/JJD/AWS
CURRENT FILING DATE: 2001-04-30
PRIOR APPLICATION NUMBER: US/09/847, 046
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: US 09/733, 288
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 807
TYPE: PRT
ORGANISM: Homo sapiens
US-09-847-046-2
Query Match 100.0%; Score 4142; DB 10; Length 807;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMSGFNFRVIGISFVIMCIFPTVNSLPELSPOKFFSTLQGLLELNEAVRPLODYGIS 60
DB 1 MMSGFNFRVIGISFVIMCIFPTVNSLPELSPOKFFSTLQGLLELNEAVRPLODYGIS 60
QY 61 VAKVVCVEEISRYGKKEKDKAVLFNGNILLREPTDTLFDVAIVAHVIFALLFSFV 120
DB 61 VAKVVCVEEISRYGKKEKDKAVLFNGNILLREPTDTLFDVAIVAHVIFALLFSFV 120
QY 121 KYTNLEDLQNIENALKKRANIIFSVYRAIGIPEHRAVMEAGFYVGTGYQFVLTTEIAL 180
DB 121 KYTNLEDLQNIENALKKRANIIFSVYRAIGIPEHRAVMEAGFYVGTGYQFVLTTEIAL 180


```

Db          441 BG--GDRLEHLSKFlE 455

RESULT 6
US-09-788-657-15
: Sequence 15, Application US/09788657
: Patent No. US20020123149A1
GENERAL INFORMATION:
APPLICANT: Nicolaiades, Nicholas
APPLICANT: Sassi, Philip
APPLICANT: Kinzler, Kenneth
APPLICANT: Grasso, Luigi
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: Methods for generating hypermutable
TITLE OF INVENTION: Yeast
FILE REFERENCE: 01107.00097
CURRENT APPLICATION NUMBER: US/09/788,657
CURRENT FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 60/184,336
PRIOR FILING DATE: 2000-02-23
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 769
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-788-657-15

```

Query Match	3.28;	Score 132;	DB 10;	Length 769;
Best Local Similarity	19.28;	Pred. No. 0.017;		
Matches 161;	Conservative 135;	Mismatches 300;	Indels 244;	Gaps 44;

```

0Y      ISAVANCVCKEISRYCGKEMDKMAATYLEKGIMILLREPTDPLFDVNAIVAHVALFALLS  118
Db      21 IISIVNNALEKMEMSIDANAFTMIDILYKEGIKVQI-TDNGSGINKRDLPI-L-CERTT   78
0Y      119 EVKYITLNEDLONEN-ALKGRANITFSYRAI-----GIPEHRAWEAGFYXGTYO---  170
Db      79 TSK-LQKFEDLSOIQTGYFGRCALASISHVAEVYTTKVEDRCAMRVASYAEGKMLESRK  137
0Y      171 -----FVLTEL-----ALLESIGSDVEYAH-----YFHCKLVLLTOOGR-  210
Db      138 PVAGKDGTTLIVEDLEFNIPSRRLRSLRNDSYKILDVYGRAIHISK---DIGSCCKFE  194
0Y      211 -----TLNEOPLTTLNIHL-----FIKTAKAPLLT-----EVAED-----POOVSTVHL--OL  251
Db      195 GDSNVSLSVKPSYTVQODIRIFYKNKSVASNLITTHISKEDNTLESVDGCNLMNFISK  254
0Y      252 GLPLVFYSQC-AIYEADRTAEWAMRLICA-----GVLLLRDSLEVNIPODANY  303
Db      255 SISLFELFNNRLVTCDLLRRALNSYSNYLPRKGFREFLYIGIVI---DPAAV----DVNV  307
0Y      304 VFKAAREGEVPVELVLDHVOLIISHVENNMHIIEEOEDDNMECPDIQODE-----VA  359
Db      308 HPTKRE-----VRFL-----SQDELETIANQJHAELSADTSRFKASSISTNKPESELIPN  360
0Y      360 ETVFRDRKRK-LPLELVELTEEFENATVMAS---DSIVLEYAQMAVSNAFLOSIDIY  414
Db      361 DTIESDRRRKSLROQOVENSYYTTANSQLCRKAKQENKLVRIDAS-QAKITSFELS--SQ  417
0Y      415 AVLKLGJTM-----LTRINCADMVDCJT-----KONYT  444
Db      418 QNFNEGSSTKROLSEPKTYTNSSHQSQAEEKLTINSEOPORDANTINDNDLKDPKKKRIK  477
0Y      445 EF--PITMYRKGENVSYAGMLGKDKLKFLOLRISYPVNTISIOEAEEYLSGLTKD  502
Db      478 DYKPSTINDERNALPISKDG-----YIRVKERYNNWLJTIKUKIREVWDJSIHRE  528
0Y      503 LI-LYSSSVYGLSPYTKTAKEDFSEAGNYLKVITIGIYSEEVDVLLSTKYASLPL  561
Db      529 LTFDFANINNYGVNDEERLAA-----IQHDKLPFLIDYGVCYEL  569
0Y      562 -----LARHTGKIESIPLASTHAQDIVOIIITDALLEMPEITIVE-----  602

```

```

Db 570 FYQIGLTDFANFGKIN---LQSTNVSD--DIVLYNLSEFDELNDDSKKEIIISKINDMS 624
QY 603 -NLPSYFRLQKPLRLIFSDGTVPNPOYKAILTLVKOKYULDSFTPCWMLKNTPYGRCILR 661
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 625 SMLNLYSYIE-----LVNDGLDND-----LKSYK-----LKSFLP---LTK 657
QY 662 AYFDPPLPLPLVLVNLHSGGVAFNPSDQAIIEENLYMLKKEAGLENIHTLTPAQEW 721
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 GYISLVLPPLPEFI---YRLKEVDWEDEQECID-----GILREIALLYIPDM 701
QY 722 KRPPLAYGYFLMIDAAATSGQTRKVPKCMKSTVDQENDKOEHDKSAVKRKEPIETLIRKH 781
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 702 VPKVDITLD-----ASLSEDKAFOFIN---KREHISL--LEH 733

```

```

RESULT 7
US-09-854-133-436
; Sequence 436, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 436
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-436

```

Query Match	3.18;	Score 129;	DB 9;	Length 396;
Best Local Similarity	20.48;	Pred. No. 0.011;		
Matches	79;	Conservative	54;	Mismatches 134;
				Indels 120;
				Gaps 15

```

QY 379 TEEFNATVMAKSDSLVEFYCAQWAVSMAFLQSYIDVAVKLG--TSMILLTRINCAWSD 436
Db 31 TADMFTHGICSAALFVEFAFWCCHQORLOPTMNDLDDKINSMEDAVYAAKVDCTAHS 90
QY 437 VCTKQNTVPEFIIMYKKGEPNVPYAGMLGTGDLKFIQJNIRISPVNITSIOAEYLS 496
Db 91 VCSAQGVAGVYPTLTKFRPGQAVAVYQGPBDQTLNNMLQTLNDEPV--TPPEVEPEPSA 148
QY 497 GELKCDLILYSVS-----VLGL-FSPYTKTAK 523
Db 149 PELQGLYEILSASFELHVAQGDHFIFAPWCCHCALAPTWEOQLALGHSBETVYICK 208
QY 524 EDFES-----AGNYLKGVITIGVSEBDVLLSTKYAASLPALLARHTEGK----- 570
Db 209 VDCIQHIELCSGNOVRY-----PTLMFR--DGKKVDDYKCK 244
QY 571 -----IES-IPLASTHADIVOJITTDALLEMPEITVENLPXYRLOKPLLLFSD 620
Db 245 RDESLREYVESQQRNETGATELVFPSEAPVLAABE-----ADKGTVALTE 293
QY 621 GTVPQYKKAALLVK-----QKYDSFPCW--LNKNTPVGRIIRAYFD----- 665
Db 294 NTFPDQTLAEGI-TEIKFYAPWCCHCKTLAPTWBELSKPEPGLAGYKIAEVDCTAERNIC 352
QY 666 ---PLPPLPLLVL-----VNLHSGCG 683
Db 353 SKYSVRCGPTLLLRGGKKVSEHSGGR 379

```

RESULT 8

PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29

PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 3.1% Score 128; DB 9; Length 432;

Best Local Similarity 20.0%; Pred. No. 0.015; Matches 77; Conservative 52; Mismatches 140; Indels 116; Gaps 14;

QY 379 TEETFNATWASDSIVLFYAGMOAVSMFLQSYIDYAVKLG--TSTMLTRINCADMSD 436
DB 67 TDMFTHGIGSAHFMFAFAPMGHCQRLOPTWMDGDKYNSMEDAKYVAVKDCJAHSD 126
QY 437 VCTKQWTEPPIITKMKKEBNPVSTAGMLGTDLKFIOLNRISTVYNITSIOEAEYIS 496
DB 127 VCSAQGVRRGYPLTKLFKPKQEAVKYQGRPDFTLENMMLQTLNEBPV--TPEVEYPPSA 184
QY 497 GELYKDLILYSYS-----VLGL-ESPMTKTAK 523
DB 185 PELKQGLYELASNEFLHVAQGDHPFKFPAPMGCHCKALAPWEOALDLEHSEYTKICK 244
QY 524 EDFSE-----AGNYLKGYVITGISEEDVLLSTKYAASLPALLARHTEGKIESIPLAS 578
DB 245 VPCQTQHYELCSGNQVGY-----PULLMR--DGKKYQYQKX 280
QY 579 THAADIQIITDALLEPFEITVENLPSYFRLOKPLLLFSDGTYNPOYRK-AITLVKQ 637
DB 281 RDESLREYVESOLORTGATEYVTPS---EAPVL-----AAPEADKGVIALTEN 330
QY 638 KYLD-----SFTPCW--LNLKNTPVGIGILRAYD----- 665
DB 331 NDDDTIAGGITPFIKTYAPWCGHCKTLAPTWBELSKKEPGLAGVIAEYDCTAERNISCK 390

OY 666 -PLPLPLVL-----VNLSGGQ 683
Db 391 YSVRGYPTLLFRGGKVSSEHSGR 415

RESULT 10

US-09-978-697-90
: Sequence 90. Application US/09978697
: Patent No. US20020169284A1
: GENERAL INFORMATION:
: APPLICANT: Ashkenazi, Avi
: APPLICANT: Baker Kevin P.
: APPLICANT: Botstein, David
: APPLICANT: Desnuyers, Luc
: APPLICANT: Eaton, Dan
: APPLICANT: Ferrara, Napoleon
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth J.
: APPLICANT: Kijavlin, Ivar J.
: APPLICANT: Kuo, Sophia S.
: APPLICANT: Napier, Mary A.
: APPLICANT: Pan, James
: APPLICANT: Paonli, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Shelton, David L.
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tunas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: P2630FIC27
: CURRENT APPLICATION NUMBER: US/09/978, 697
: CURRENT FILING DATE: 2001-10-16
: PRIOR APPLICATION NUMBER: 09/918585
: PRIOR FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/064249
: PRIOR FILING DATE: 1997-11-03
: PRIOR APPLICATION NUMBER: 60/065311
: PRIOR FILING DATE: 1997-11-13
: PRIOR APPLICATION NUMBER: 60/066364
: PRIOR FILING DATE: 1997-11-21
: PRIOR APPLICATION NUMBER: 60/077450
: PRIOR FILING DATE: 1998-03-10
: PRIOR APPLICATION NUMBER: 60/077632
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077641
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077649
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077791
: PRIOR FILING DATE: 1998-03-12
: PRIOR APPLICATION NUMBER: 60/078004
: PRIOR FILING DATE: 1998-03-13
: PRIOR APPLICATION NUMBER: 60/078886
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078936
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078939
: PRIOR FILING DATE: 1998-03-20

: PRIOR APPLICATION NUMBER: 60/079294
: PRIOR FILING DATE: 1998-03-25
: PRIOR APPLICATION NUMBER: 60/079656
: PRIOR FILING DATE: 1998-03-26
: PRIOR APPLICATION NUMBER: 60/079664
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079663
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079728
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079786
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079920
: PRIOR FILING DATE: 1998-03-30
: PRIOR APPLICATION NUMBER: 60/079923
: PRIOR FILING DATE: 1998-03-30
: PRIOR APPLICATION NUMBER: 60/080105
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080107
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080165
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080194
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080327
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/080328
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/080333
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/080334
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/081070
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081049
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081071
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081195
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081203
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081229
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081955
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081817
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081819
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081952
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081838
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/082568
: PRIOR FILING DATE: 1998-04-21
: PRIOR APPLICATION NUMBER: 60/082569
: PRIOR FILING DATE: 1998-04-21
: PRIOR APPLICATION NUMBER: 60/082704
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082804
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082700
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082797
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082796
: PRIOR FILING DATE: 1998-04-23
: PRIOR APPLICATION NUMBER: 60/083336
: PRIOR FILING DATE: 1998-04-27
: PRIOR APPLICATION NUMBER: 60/083322

```

1 PRIOR FILING DATE: 1998-04-28
2 PRIOR APPLICATION NUMBER: 60/083392
3 PRIOR FILING DATE: 1998-04-29
4 PRIOR APPLICATION NUMBER: 60/083495
5 PRIOR FILING DATE: 1998-04-29
6 PRIOR APPLICATION NUMBER: 60/083496
7 PRIOR FILING DATE: 1998-04-29
8 PRIOR APPLICATION NUMBER: 60/083499
9 PRIOR FILING DATE: 1998-04-29
10 PRIOR APPLICATION NUMBER: 60/083545
11 PRIOR FILING DATE: 1998-04-29
12 PRIOR APPLICATION NUMBER: 60/083554
13 PRIOR FILING DATE: 1998-04-29
14 PRIOR APPLICATION NUMBER: 60/083558
15 PRIOR FILING DATE: 1998-04-29
16 PRIOR APPLICATION NUMBER: 60/083559
17 PRIOR FILING DATE: 1998-04-29
18 PRIOR APPLICATION NUMBER: 60/083500
19 PRIOR FILING DATE: 1998-04-29
20 PRIOR APPLICATION NUMBER: 60/083742
21 PRIOR FILING DATE: 1998-04-30
22 PRIOR APPLICATION NUMBER: 60/084366
23 PRIOR FILING DATE: 1998-05-05
24 PRIOR APPLICATION NUMBER: 60/084414
25 PRIOR FILING DATE: 1998-05-06
26 PRIOR APPLICATION NUMBER: 60/084441
27 PRIOR FILING DATE: 1998-05-06
28 PRIOR APPLICATION NUMBER: 60/084637
29 PRIOR FILING DATE: 1998-05-07
30 PRIOR APPLICATION NUMBER: 60/084639
31 PRIOR FILING DATE: 1998-05-07
32 PRIOR APPLICATION NUMBER: 60/084640
33 PRIOR FILING DATE: 1998-05-07
34 PRIOR APPLICATION NUMBER: 60/084598
35 PRIOR FILING DATE: 1998-05-07
36 PRIOR APPLICATION NUMBER: 60/084600
37 PRIOR FILING DATE: 1998-05-07
38 PRIOR APPLICATION NUMBER: 60/084627
39 PRIOR FILING DATE: 1998-05-07
40 PRIOR APPLICATION NUMBER: 60/084643
41 PRIOR FILING DATE: 1998-05-07
42 PRIOR APPLICATION NUMBER: 60/085339
43 PRIOR FILING DATE: 1998-05-13
44 PRIOR APPLICATION NUMBER: 60/085338
45 PRIOR FILING DATE: 1998-05-13
46 PRIOR APPLICATION NUMBER: 60/085323
47 PRIOR FILING DATE: 1998-05-13
48 PRIOR APPLICATION NUMBER: 60/085582
49 PRIOR FILING DATE: 1998-05-15
50 PRIOR APPLICATION NUMBER: 60/085700
51 PRIOR FILING DATE: 1998-05-15
52 PRIOR APPLICATION NUMBER: 60/085689
53 PRIOR FILING DATE: 1998-05-15
54 PRIOR APPLICATION NUMBER: 60/085579
55 PRIOR FILING DATE: 1998-05-15
56 PRIOR APPLICATION NUMBER: 60/085580
57 PRIOR FILING DATE: 1998-05-15
58 PRIOR APPLICATION NUMBER: 60/085573
59 PRIOR FILING DATE: 1998-05-15
60 PRIOR APPLICATION NUMBER: 60/085704
61 PRIOR FILING DATE: 1998-05-15
62 PRIOR APPLICATION NUMBER: 60/085697
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

Db	127	VCSAQGVGRVPTLLFKFPGCAAVVAGYGRPRQGLTLENNMLQTLEENPY--	-TPPEEVEEPPSA	188
QY	497	GELYKDLLIVSSVS-----	-VLGL-FSPPTMTAK	523
Db	185	PELKQGLYEVSASFELHVAAGDHFIKFPAPWGCHKALAPTWBQALSGHSHYATIGK	244	
QY	524	EDFSE-----AGNYLKGVITIGIYSEEDVLLSTKYAASLPALLARHTEGKTESIPLAS	578	
Db	245	VDCTQHLELCSGQNVGRY-----	PTLLMFR--GDKKVDQYKKG	280
QY	579	THAQDIQIITDALLEFPELTIVENLSPYERLQKPLILFSDGVNVQYKK-AIITLVKQ	637	
Db	281	RDELSELRVEYSQLORETGATETVTS-----EAPVY-----AAPEADKGTVALTEN	330	
QY	638	KYLD-----	SFPFCM--INLKNTPYVGRGLIRAYFD-----	665
Db	331	NFDDTIAEGITFIKFPAPWGCHKTLAPTWBELSKKEFPGLAGVKIALEVDTAEARNICK	390	
QY	666	PLPPLPLVY-----VNLHSGQ	683	
Db	391	YSVAGVPTLLLFGRGKRVKSEHSGGR	415	

RESULT 11
 US-09-978-90
 : Sequence 90, Application US/09978192A
 : Patent No. US20020177553A1
 : GENERAL INFORMATION:
 : APPLICANT: Ashkenazi, Avi
 : APPLICANT: Baker Kevin P.
 : APPLICANT: Botstein, David
 : APPLICANT: Desnoyers, Luc
 : APPLICANT: Eaton, Dan
 : APPLICANT: Ferrara, Napoleon
 : APPLICANT: Filvaroff, Ellen
 : APPLICANT: Fong, Sherman
 : APPLICANT: Gao, Wei-Qiang
 : APPLICANT: Gerber, Hanspeter
 : APPLICANT: Gerltsen, Mary E.
 : APPLICANT: Goddard, Audrey
 : APPLICANT: Godowski, Paul J.
 : APPLICANT: Grimaldi, J. Christopher
 : APPLICANT: Gurney, Austin L.
 : APPLICANT: Hillan, Kenneth J.
 : APPLICANT: Kijavlin, Ivar J.
 : APPLICANT: Kuo, Sophia S.
 : APPLICANT: Napier, Mary A.
 : APPLICANT: Pan, James J.
 : APPLICANT: Paoni, Nicholas F.
 : APPLICANT: Roy, Margaret Ann
 : APPLICANT: Shelton, David L.
 : APPLICANT: Stewart, Timothy A.
 : APPLICANT: Tumas, Daniel
 : APPLICANT: Williams, P. Mickey
 : APPLICANT: Wood, William I.
 : TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 : FILE REFERENCE: P2630P1C9
 : CURRENT APPLICATION NUMBER: US/09/978,192A
 : CURRENT FILING DATE: 2001-10-15
 : PRIOR APPLICATION NUMBER: 09/918585
 : PRIOR FILING DATE: 2001-07-30
 : PRIOR APPLICATION NUMBER: 60/062250
 : PRIOR FILING DATE: 1997-10-17
 : PRIOR APPLICATION NUMBER: 60/064249
 : PRIOR FILING DATE: 1997-11-03
 : PRIOR APPLICATION NUMBER: 60/065311
 : PRIOR FILING DATE: 1997-11-13
 : PRIOR APPLICATION NUMBER: 60/066364
 : PRIOR FILING DATE: 1997-11-21
 : PRIOR APPLICATION NUMBER: 60/077450
 : PRIOR FILING DATE: 1998-03-10
 : PRIOR APPLICATION NUMBER: 60/077632


```
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match      3.1%; Score 128; DB 9; Length 432;
Best Local Similarity 20.0%; Pred. No. 0.015;
Matches 77; Conservative 52; Mismatches 140; Indels 116; Gaps 14;

QY 379 TEETFNATVVASDSIVLFYFGMAVSNAPFQSIYIDVAVKLG--TSMMLTRINCADMSD 436
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 67 TADMFTHGIDSAHFWEFFAPWCGHCGRLOPTWMDLDDKNSMEDAKVYAKVCTAHSD 126
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 437 VCTKQNTPEPTIKMYKKGPNVSYAGMLGSKDLEKFIQINRISYPNITSIOAEYLS 496
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 127 VCSAQGYRGPTLKLFRPGGEAVKYGCPROFQILENNMLOTLNEEPV--TPPEVEPPSA 184
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 497 GELYKDLILYSVS-----VLGL-FSPPTMTAK 523
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 185 PELKQGLYELASNFELHVAQGDHFIFAPWCGHCKALAPWEOQLALGLEHSEYVIGK 244
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 524 EDSE-----AGNYLKGIVTGIYSEEDVLLSTKYASLPAILLANHTGKIESIPLAS 578
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 245 VDCQHYELCSGNQVRGY-----PTLLWFR--DGKKVDQYK 280
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 579 THADQIVQITDALLEFPEITVENLPSYPLQKPLILFSDGVNPOYK-AIITLVKQ 637
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 281 RQLESLEREYVESQLORETCATETVPS---EAPVL-----AAPEAKGVIALTEN 330
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 638 KYLD-----SFTPCN--LNKNTPVYRGILRAYFD----- 665
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 331 NFDDTIAEGITFTKFPYAPWCGHCKTLAPWELSKKEFPGLAGVIAEVCTAERNICSK 390
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 666 -PLPLPLVL-----VNLHSGQ 683
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 391 YSVRGYPTLLFRGKKVSEHSGR 415
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12
US-09-999-832A-90
; Sequence 90, Application US/09999832A
; Publication No. US20020192706A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
```

```
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630PIC3
; CURRENT APPLICATION NUMBER: US/09/999, 832A
; CURRENT FILING DATE: 2001-10-24
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
```

```

; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081952
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082700
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082796
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: 60/083336
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083392
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083558
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083500
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083742
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07

; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match          3.1%; Score 128; DB 9; Length 432;
Best Local Similarity 20.0%; Pred. No. 0.015;
Matches 77; Conservative 52; Mismatches 140; Indels 116; Gaps 14;

OY 379 TETFNATVWASDSIYFYAGWQAVSMAPLOSIDYAVKLG--ISTMLITRINCADMSD 436
    | : : : : : | : : : : : | : : : : : | : : : : : |
DB 67 TADMFTHGIGIOSAAHFVMPFAPMCGHCQRLQPTWINDLGDKYNSMEDAKVYAVKVDCTAHS 126

OY 437 VCTKQNVTEFFIITIKWKKEPNVSYAGMGTDLKFIOLNISTVPVNTISQDEAEYIS 496
    | : : : : : | : : : : : | : : : : : | : : : : : |
DB 127 VCSAGVRCYPTLKLKPEQEAVKYQGPDPFQTLLENMMLQTLNEEV--TPPEVEPPSA 184

OY 497 GELYKDLILYSVS-----VLGL-ESPWTAK 523
    | : : : : : | : : : : : | : : : : : | : : : : : |
DB 185 PELKQGLYELSSNPELHNAOGDHFIFAPMCGCHKALAPWEOALGLEHSEYIK 244

OY 524 EDFSE-----AGNYLKGYITGYSEEDVLLSTKYAASLPALLARHTEGIESIPLAS 578
    | : : : : : | : : : : : | : : : : : | : : : : : |
DB 245 VCTQHYELCSGNQVGY-----PILMER--DGKKVQYK 280

OY 579 THAODIVQITPALLEMPFEITVENIPSYRLOKPLILIFSQYVNPQYK-AIITLVQ 637
    | : : : : : | : : : : : | : : : : : | : : : : : |
DB 281 RLESILREYVESOLQRTETGATEYTPS---EAPVL-----AAPEADKGTVALTEN 330

OY 638 KYLD-----SFTPCM--LNLKNTPVGRGILRAYPD----- 665
    | : : : : : | : : : : : | : : : : : | : : : : : |
DB 331 NFDDTIAEGITFIKYFAPMCGCHKTLAPWELSKKEFFGLAGVAKIAYDCTAERNICK 390

OY 666 -PLPPLILVL-----VNLHSGQ 683
    | : : : : : | : : : : : | : : : : : | : : : : : |
DB 391 YSVRGVPTLLFRGKGVSEHSGR 415

RESULT 13
US-09-978-189-90
; Sequence 90, Application US/09978189
; Publication No. US20030004102A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botsstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerltsen, Mary E.
```

APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James;
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC7
CURRENT APPLICATION NUMBER: US/09/978,109
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366

APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P34301C104
CURRENT APPLICATION NUMBER: US/10/176,758
CURRENT FILING DATE: 2002-06-21
Prior Application removed - See file wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 74
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-758-74

Query Match 3.18; Score 128; DB 9; Length 432;

Best Local Similarity 20.0%; Pred. No. 0.015;

Matches 77; Conservative 52; Mismatches 140; Indels 116; Gaps 14;

QY 379 TEEFNATVWASDSIVFYWGQAVSMATFQSYIDVAVKLG--TSMLTRINCAWDSD 436
DB 67 TADFTHTGIGSAHFVFFAFPCGHCORLOPTMNDLGDKNMEDAKYVAKVDCTAHD 126
QY 437 VCTKQNTTEPPIKMYKKGPNVSYAGMLGTDLKEIOLNRISYPVNITSIQEAEYLS 496
DB 127 VCSAQGVGRVPTLKLKFGQEAHVYQGRDQGTLENMLOTLINEPV--TPEPEVEPPSA 184
QY 497 GELYKDLIXSVS-----VLGL-FSPMTAK 523
DB 185 PELKQGLYELASNFELHVAQGHFIKFPAPWCCHCKALAPWEOALGLEHSETVKIGK 244
QY 524 EDESE-----AGNYLKGYVTGTIYSEEDVLLSTKYAASLPALLARHTEGKTESIPLAS 578
DB 245 VDCTQHIELCSGNQVRY-----PTLMFR--DGKKVDQYK 280
QY 579 THADIVQITTDALLEMEPEITVENLPSYFRLQPLILFSDGTVPQYK-ALLFLVKQ 637
DB 281 RLESLEREYESQLORETGATETVPS---EAPVL-----AAEPADKGTVALTEN 330
QY 638 KYLD-----SFTPCW--LNLKNTPVGRGILLRAYFD----- 665
DB 331 NFDDTIAEGITFTKFPAPWCCHCKTLAPTWELSKKEFPGLAGYKIAEVDCTAERNICSK 390
QY 666 -PLPLPLVL-----VNLHSGGQ 683
DB 391 YSVRGYPTLLFRGKKVSEHSGR 415

Search completed: March 13, 2003, 17:18:06
Job time : 21 secs

[illegible]

RESULT 2
P96562
hypothetical protein F19K6.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: P96562
R:Phlogiotis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitli, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712
A:Accession: P96562
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-546 <STO>
A:Cross-references: GB:AE005173; NID:g10645454; PIDN:ANG21570.1; GSPDB:GN00141
C:genetics:
A:Gene: F19K6.17
A:Map position: 1

Query Match	4.4%;	Score 182;	DB 2;	Length 546;
Best Local Similarity	20.9%;	Pred. No. 0.00041;		
Matches 111;	Conservative 93;	Mismatches 180;	Indels 148;	Gaps 25

```

OY 317 LVLMHVDDLLSIVENNNHHEIOEDEDNDME---GPDIDVOOD-----EVAETVRODKR 368
Db 19 LLLLSFTIIITIAVSSPDSNVESNEPFGSDLDLOLLAVDLOEDPREQOSEATV--SKAQ 76
OY 369 KLPLELVEELTEETFNATVMSDSIVLEFYAGWQVMSAFLOSVIDAVKLGKT-STMLLT 427
Db 77 RIVLELNGDYTKRAVIDGNEFV---MVLGYAPWCARSALMPPFAAATAVALNEIGSSVIMA 133
OY 428 RINCADMSDVCTKONATEFPITIKMKKGENPVSYAGMLGTDLKFLQTOLNRIISYVNTS 487
Db 134 KIDGDRYSKIASIELEIKGFPTLLLEFVNGTS--LTYNGGSSAEDIVIVOKKGAPIITLNT 192
OY 488 IOBAEYVLSGELYKLLIYSSVYGLF-----SPTMKTAKEDE--FSE----- 528
Db 193 VDEAPRFL--DKYHTF-----VLGLEKEFGSHNEFVKAASDDEIOFLERDSDVA 243
OY 529 -----AGNYLKGY-----VIRGIYSEEDVL--LLSTKY----- 554
Db 244 KLPLPDLKSNVFIGLVAKPEAERYTVYDGSYKMEKILEFLOSNNKFLPTKLTLENTVAVY 303
OY 555 --AASLPALLLARHTEGKIESIPLASTHAODIVQIITDALLEMPETIVENLPSTYRLQK 612
Db 304 SSPVYLQWMLFSKADDFOKLAQPL-----EDIAKFKSKLMFIYDVITNEN-----DAM 352
OY 613 PLTILFSGGYNVPOYKKAAILTVLVKOKYLDSTFCMLNKN--TPVGRILRAYF--DLP 668
Db 353 PLLILFGEIAGNKTVAAFDNNNLNSKYLLIESDPSPNLSIEEPCSGIAHGTATRYVRSSEVP 412

```

```

Oy 669 PLPLIVLVNLSGOVFAPFDQITII---EELVYV-----PLTKLEIA-----G 708
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 DNEANSIYV---YKRTF---DGLVYLSRENVLELVHTPVCYVNCALSKQIEKLAKHPG 466
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 709 LFNHTITLPAQEMKPLPAIDFLSMDIAASQKSTKRVPCMKMETVQOENDK 760
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 467 FENLV-----FRIDASANE-----HTKIQVNDK 490
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 3
S44756
probable protein disulfide-isomerase (EC 5.3.4.1) - *Caenorhabditis elegans*
N:Alternate names: C14B9.2 protein
C:Species: *Caenorhabditis elegans*
C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 16-Jul-1999
C:Accession: S44756
R:Favell, A.D. *Dev Biol* 147:100-107, 1992

submitted to the EMBL data library, May 1993
A:Description: Sequence of the C. elegans cosmid C14B9.
A:Reference number: S44617
A:Accession: S44756
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-664 <FAV>
A:Cross-references: EMBL:L15188; NID:g289640; PIDD:AAA27952.1; PID:g289643
C:Genetics:
A:Introns: 43/2; 110/2; 631/3
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: intramolecular oxidoreductase; isomerase
E:90-170/Domain: thioredoxin homology <TX1>
E:201-285/Domain: thioredoxin homology <TX2>
E:553-640/Domain: thioredoxin homology <TX3>

Query Match	4.2%;	Score 172.5;	DB 2;	Length 664;
Best Local Similarity	19.4%;	Pred. No. 0.0024;		
Matches 87;	Conservative 90;	Mismatches 175;	Indels 97;	Gaps 17;

```

QY 345 DMEBP-DIDVODDE--VAENFYFR--DGRKRLPLELTYLLEFNFNNTVMASSIVL--FYA 398
Db 160 DKGKGNDDGGRDAGIIVEMESVVDNYKPPREBYVLTTFENFDOPISNNELVLEFYTA 219
QY 399 GWOAVSMAFLQSYIDVAVNVLKGT-STMLTRINACADMSDVCQONVTEPILIKMYKGEN 457
Db 220 PWCCHCKKLAPERYEKAQKILQAQSSKVLKGVADATTEKDGTYGVSGPYTKIIRNGRR 279
QY 458 PVSYAGMLGTDKLKFQOLNRISPVNVTISIQBAEELTSELKDLILYSSVVLGLESFP 517
Db 280 -FDYNGPREAGIILIKYMTDQSKPAKKLPRKLDOVERMS---KD---DVTIIGEFAT 329
QY 518 TMTAKKEDFSEAGNYLKGVIYITGYSSE-----DVLLSTKYAASLPALLLR 565
Db 330 EDSTAFEFESDSAMLLEEFKTMCHTSDPAFKKMDAKPRDIIIF-----YPSLFHFK 382
QY 566 HTECKIESIPLASTHAODIVOIITDALLEPPELTYENLPSYFRLOKPLLILFSDGTVP 625
Db 383 F-EKRSTYTNKAAATSEDLAFREHSAPLVGMKTRKNATRY-TKKPLVAVYVYNADFVSU 440
QY 626 QYKKA-----ILTLVKOKYLDSEFPQCMLEKNTFVGRIILAVYDPLPLPLVLVNL 678
Db 441 QYRGSESYWMSKYVNLNIOKQIOKDKYK----- 466
QY 679 HSGGVFAFPSSDQIIEENLVMLKLK--EAGLENNHITILPAQEWKPRPLPAVDFLSMID 735
Db 467 -----FAVDADEEFAKE-----LEELGLDGLSGLEHNHVVGVGDYKKYPMNPDEFGEILD 515
QY 736 -----AATSQKTRKYPRKCMETVQOENDK 760
Db 516 ENLEAFKQIJSKGAKAHVNSAPRKKDK 544

```

RESULT 4
S34275
protein disulfide-isomerase homolog precursor - fluke (Schistosoma mansoni)

A:Accession: A41713
A:Molecule type: DNA
A:Residues: 1-82, 'V', 84-192, 'S', 144, 146-167, 'E', 169-225, 'V', 227-457, 'S', 459-505, 'EADAEAE'
A:Cross-references: GB:M76992; NID:g173023; PIDN:AAA35169.1; PID:g173024
R:Dubois, E.; Pletard, A.; Gligol, D.; Glansdorff, N.; Messenguy, F.; Scherrens, B.
Submitted to the Protein Sequence Database, March 1992
A:Reference number: S19367
A:Accession: S19372
A:Molecule type: DNA
A:Residues: 1-522 <DUB>
A:Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA44273.1; PID:g55320; GSPDB:GN00003
R:Scherrens, B.; Messenguy, F.; Gligol, D.; Dubois, E.
Least 8, 577-586, 1992
A:Title: The complete sequence of a 9,543 bp segment on the left arm of chromosome III
A:Reference number: S25347; MUID:92397595; PMID:1523890
A:Accession: S25349
A:Molecule type: DNA
A:Residues: 1-522 <SG2>
A:Cross-references: EMBL:X59720; EMBL:S43845; EMBL:S49180; EMBL:S58084; EMBL:S93798; NIT
A:Genetics:
A:Gene: SGD:PD11; MPP1, TRG1, MIPS:YCL043C
A:Cross-references: SGD:S0000548; MIPS:YCL043C
A:Map position: 3L
C:Function:
A:Description: catalyzes rearrangement of both intrachain and interchain disulfide bonds
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
E:keyWords: duplication; endoplasmic reticulum; glycoprotein; homodimer; intramolecular
E:1-22/Domain: signal sequence #status predicted <Sig>
E:23-522/Product: protein disulfide isomerase #status predicted <Sig>
E:40-125/Domain: thioredoxin homology <TX2>
E:384-470/Domain: thioredoxin homology <TX2>
E:519-522/Region: endoplasmic reticulum retention signal
E:61-64, 406-409/Disulfide bonds: redox-active #status predicted
E:82, 117, 155, 174, 425/Binding site: catalytic site (Asn) (covalent) #status predicted
E:90-97/Disulfide bonds: #status predicted

```

Query Match          3.7%; Score 154.5; DB 1; Length 522;
Best Local Similarity 20.88; Pred. No. 0.025;
Matches 66; Conservative 67; Mismatches 126; Indels 58; Gaps 15;

QY 358 VAETVFDRKRRLPLE-LTVELTEETFNATVMSDSIVL-FYAGQAVSMALQSYIDVA 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LASSVFAQGEAVAPEDSAVVKLATDSFNEXIQSHDLVLAEPFAPMGCHCKNMADVEYKAA 75

QY 416 VKLKGSTMLTRINCADWDVCTKQNTPEPIIKMKKGE--NPVSAGMLGTRDLIKE 473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 ETLV-EKNITLAQIDCTENODLCEMHNIPGPSKIFENSDVNNISIDEGERTEAIVQF 134

QY 474 IQLNRISPYNVITSIOEAEITSGELYKDLILYSVSVLGFSPTMKAKEDFEPAGNYL 533
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 --MIKQSPQ-VAAYVADIPATLANETFTVPVIYVSGKIDADFNNATF-----YSMANKHF 185

QY 534 KGYVITGTISEEDVLLSTKYAASL--PALLLARHTEGKIESIPLASTHAODIYQIITDA 591
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 NDYQFVSAENMDDDFKLSITVLPMSADEPVYVNGKKAD-----IADA 226

QY 592 LLEMFPE-IIVENLPSFTRLOKPLLIFFSGDGVNPQYKKAILLT-----VKQKLYDSETP 645
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 227 --DYFEKWLQVEALPYFGEI-----DSGVFAQYVESGLPLGLVLFYNDEEELAEYKP 275

QY 646 CWLNL--KNTPVGRGIL 660
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 LFTELAKKN---RGLM 288

RESULT 12
A28807
protein dlsulfide-isomerase (EC 5.3.4.1) ER60 precursor - rat
N:Alternate names: endoplasmic reticulum protein ERp60; ER60 proteinase; hormone induced
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Mar-1989 #sequence,revision 27-Jun-1994 #text_change 18-Jun-1995
C:Accession: A28807; A43050; A43051; A40095
C:Bennett, C.F.; Balcarek, J.M.; Varrichio, A.; Crooke, S.T.

```

Nature 334, 268-270, 1988
A>Title: Molecular cloning and complete amino-acid sequence of form-I phosphoinositide
A:Reference number: A28807; MUID:88288403; PMID:3398923
A:Accession: A28807
A:Molecule type: RNA
A:Residues: 1-504 <EN>
A:Cross-references: GB:112355; NID:956904; PIDN:CAA0916.1; PID:956905
R:Martin, J.L.; Punford, N.R.; Laposo, A.C.; Martin, B.M.; Gonzaga, H.M.S.; Beaven, M
Biochem. Biophys. Res. Commun. 178, 679-685, 1991
A>Title: A metabolite of halothane covalently binds to an endoplasmic reticulum prote
A:Reference number: A43050; MUID:91315499; PMID:1650195
A:Accession: A43050
A:Molecule type: protein
A:Residues: 25-37/95-104/130-138/274-279/366-378/426-450/482-495 <AR>
R:Urade, R.; Nasu, M.; Moriyama, T.; Wada, K.; Kito, M.
J. Biol. Chem. 267, 15152-15159, 1992
A>Title: Protein degradation by the phosphoinositide-specific phospholipase C-alpha f
A:Reference number: A43051; MUID:92340568; PMID:121829
A:Accession: A43051
A:Molecule type: protein
A:Residues: 25-34/173-194/432-459 <UR>
R:Mobbs, C.V.; Fink, G.; Platt, D.W.
Science 247, 1477-1479, 1990
A>Title: RHP-10: a protein induced by estrogen in the brain and LH-RH in the pituitary
A:Reference number: A40095; MUID:90208308; PMID:2181662
A:Accession: A40095
A:Molecule type: protein
A:Residues: 26-43 <MB>
C:Comment: ER60 may be part of a complex capable of catalyzing cleavage of itself, of
C:Comment: ER60 protease does not appear to be a cytoplasmic 1-phosphatidylinositol-4
inositol-4-phosphate, or phosphatidylinositol-4,6-bisphosphate.
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: duplication; endoplasmic reticulum; intramolecular oxidoreductase; isomer
F:1-24/Domain: signal sequence #status predicted <Sig>
F:25-504/Product: ER60 protease #status experimental <MA>
F:33-118/Domain: thioredoxin homology <TX1>
F:383-469/Domain: thioredoxin homology <TX2>
F:501-504/Region: endoplasmic reticulum retention signal #status atypical
F:57-60/405-408/Dsulfide bonds: redox-active #status predicted

[illegible]

Arch. Biochem. Biophys. 308, 454-460, 1994
A:Title: Ekp61 is Grp58, a stress-inducible luminal endoplasmic reticulum protein, but is
A:Reference number: S41661; MUID:94153092; PMID:8109975
A:Accession: S41661
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-504 <MA>
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: endoplasmic reticulum; intramolecular oxidoreductase; isomerase; redox-activ
F:33-118/Domain: thioredoxin homology <TX1>
F:383-469/Domain: thioredoxin homology <TX2>

Query Match 3.7%; Score 152; DB 2; Length 504;
Best Local Similarity 20.7%; Pred. No. 0.034;
Matches 61; Conservative 53; Mismatches 123; Indels 58; Gaps 9;

Qy 376 VELTEFNATYMASDS----VLFYAGQAVSNAPFLOSIDYAVKLGKSTMLTRINC 431
Db 28 LELVDENESRVSDFGAGMLVEFFAPWCCHKRLAPEYEAATRLKG--TVPLAKYDC 85
Qy 432 ADMSPVCTKONVTEPPIIKMYKKGPNVSYAGMLGKTKDLKFIOLNRISYPVNTSIOEA 491
Db 86 TANTWTCKKYGVTGYPTLIKIFPDGEAGAYDGPRTADGIVS--HLKKQAGPASYPLRTED 143
Qy 492 EBYLSGELYKDLIYSSVVLGLFSPYTKTAKEDSEAGNYLKQYV-----ITGIYSEE 545
Db 144 E-----FKKFISDKDASVVGFFRDLPDSGHSEFLKAASNLRDYVRFATHTVESLVKEY 196
Qy 546 D-----VLLSTYVNASLPLALLARHTGKIEST-----PLASTHAODIYQI 587
Db 197 DONGGRIIFRPLHLANKFEEDKIVAYTEKKMTSGKSRSIFRKAFCGLCPNTEEDNKDLIO- 255
Qy 588 ITDALLEMPETLVNLPSEYRLOKPLILFSDGTVPYQYKAILTVLKQKILDS 642
Db 256 -----GKDLTAIYDIDV-----YKNTKGSNMYNRNRYMAAKFTLDA 292

RESULT 14

JC7623
protein disulfide-isomerase (EC 5.3.4.1) precursor - yeast (*Pichia pastoris*)
C:Species: *Pichia pastoris*
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 24-Aug-2001
C:Accession: JC7623
R:Masame, A.; Vad, R.; Kristensen, T.; Oyen, T.B.
Biochem. Biophys. Res. Commun. 281, 1176-1182, 2001
A:Title: Characterization of a gene encoding a *Pichia pastoris* protein disulfide isomera
A:Reference number: JC7623; MUID:21139750; PMID:11243858
A:Accession: JC7623
A:Molecule type: DNA
A:Residues: 1-517 <MA>
A:Cross-references: GB:AJ307014
A:Experimental source: strain GS115
C:Genetics:
A:Gene: pdi
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: intramolecular oxidoreductase; isomerase; multifunctional enzyme
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-517/Product: protein disulfide-isomerase #status predicted <MAT>
F:59-66/Domain: redox-active #status predicted <REAI>
F:401-408/Domain: redox-active #status predicted <REAI>
F:514-517/Region: endoplasmic reticulum retrieval signal #status predicted

Query Match 3.6%; Score 150.5; DB 2; Length 517;
Best Local Similarity 22.5%; Pred. No. 0.044;
Matches 79; Conservative 64; Mismatches 161; Indels 47; Gaps 15;

Qy 376 VELTEFNATYMASDSIVL--FYAGQAVSNAPFLOSIDYAVKLGKSTMLTRINCADW 434
Db 36 VKLTFATFESFTSNPHVLAEPFAPWCCHKKLGPELVSAAILDNDEQVLAQIDCTEE 95
Qy 435 SDVCTKONVTEPPIIKMYK--KGENPVSYAGMLGKTKDLKFIOLNRISYPVNTSIOEAE 493
Db 96 KELCGYEIKGYPTLVKFGVEVPSDYQGQROSOSIYSYMLKQSLPVPSEINATKDLDD 155

Qy 494 YLSGELYKDLIYSSVVLGLFSPYTKTAKEDSEAGNYLKQYITGIYSEE--DVLLIS 551
Db 156 TIA-EAKEPVI-----YQVL-----PEASNLESNTTGYAGTLREKFTFYSTS 200
Qy 552 TRYAA-----SLPALLARHTGKIEST-----PLASTHAODIYQIITDALMPETIVE 602
Db 201 TTYAKKRYTSDSTNPAYLVLR--PGEEPSVSGELEDETHLVHWIDIESKPL--FGDIDGS 255
Qy 603 NLPSEYRLOKPLILFSDGTVPYQYKAILTVK-----QKLDSTFPCWMLNKTPVGR 657
Db 256 TFKSYAEANIPLAYYYFE--NEQRAAAADIIKPPAKDQKKNF---VGLDAVKFK 308
Qy 658 GILRAYFDPLPPLVNLVNLHSGGVFAFPDQAIIEENLVMLKLEAG 708
Db 309 HAKNINMDE-EKLPLEVHIDLVS--KKKFGVPDQDELTKNDVIELLEKFIAG 357

RESULT 15

T06724
protein disulfide-isomerase homolog F28P10.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06724
R:Queller, F.; Choisme, N.; Robert, C.; Brotlier, P.; Wincker, P.; Catolico, L.; Art
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15793
A:Accession: T06724
A:Molecule type: DNA
A:Residues: 1-566 <QDE>
A:Cross-references: EMBL:AL049655; GSPDB:GN00061; ATSP:F28P10.60
A:Experimental source: cultivar Columbia; BAC clone F28P10
C:Genetics:
A:Gene: ATSP:F28P10.60
A:Map position: 3
A:Introns: 192/2; 235/3; 291/1; 330/3; 351/3; 380/3; 405/3; 428/3; 455/3; 493/3; 512/
C:Superfamily: protein disulfide-isomerase; thioredoxin homology

Query Match 3.6%; Score 148; DB 2; Length 566;
Best Local Similarity 20.5%; Pred. No. 0.074;
Matches 129; Conservative 71; Mismatches 200; Indels 228; Gaps 28;

Qy 286 LILLRDSLEVNIPQ-----DANVVFKAEEGVVEEFLVLDVNLISHVNNHIEIORD 341
Db 17 ILLVNSRRENNSSGSDDELAFLAEESEKQ-----SHGGSYH-EEENH 63
Qy 342 EDNDMEG-----PDIVDDVEAETVFRDRRKLPLETVEL 378
Db 64 QHRDFENVDDLQGGGEFHNGDHYEEEPPLPVDEKDAV-----L 104
Qy 379 TREETNATY--MASDSIVL--FYAGQAVSNAPFLOSIDYAVKLGKSTMLTRINCADMSV 437
Db 105 TKDNFTFVGNNSFAMVEEYAPWCACQALTPEDYAAATLKLGLAA--LAKIDATEEGDL 162
Qy 438 CTKONVTEPPIIKMYKKGPNVSYAGMLGKTKDLKFIOLNRISYPVNTSIOEAEYISG 497
Db 163 AQKYLEGPGPYFLFVVDGEMRKRTYEGEKDKDQIVMLKKKASPISHNITTKREAEVLSA 222
Qy 498 E-----LYKDLIYSSVS--VLGLF-----PYMKTAKE 524
Db 223 EKLVLGFLNLSVGSSEELAASRLDLSFYQTASPDIARLFETIQVKRPAYLVLLK 282
Qy 525 DFSEA-----GNVLKGVITGIYSEEDVLLS--TKYAA--LPALL 562
Db 283 EEEKLAREDNFTKTAIAEFVSAKNVPVIVNFTBQASLIFSSVKNQANSESEKHLPTLR 342
Qy 563 -LARTTEGKI-----ESIP-----LASTHAODIYQIITDALLEMF 596
Db 343 EVAKSFKGFVYVYQOMNEDYGEAVSGFVGTAAPKLVLYTGNDMKKFLIDG----- 397
Qy 597 PEITVENLPSYRLOKPLILFSDGTVPQYK-----KAI-----LTLVKQ 637
Db 398 -ELTVNNI-----KTLAEFLADKLKPFYKSDPLPENNDGVKVIYGNNEDEIVLDES 449

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:10:57 ; Search time 16 Seconds

(without alignments)
2091.962 Million cell updates/sec

Title: US-09-847-046-2

Perfect score: 4142
Sequence: 1 MFSGFNFRVIGISFVIMCIF.....FKEAKSPRRDKELGCSKVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	185	4.5	505	PDI_HUMAN	P55059 humicola in
2	177	4.3	496	PDI_DROME	P54399 drosophila
3	175.5	4.2	505	PDA3_BOVIN	P38657 bos taurus
4	172.5	4.2	515	PDI_ASPOR	Q00248 aspergillus
5	172.5	4.2	618	PDA4_CAELT	P34329 caenorhabdi
6	169.5	4.1	505	PDA3_HUMAN	P30101 homo sapien
7	169.5	4.1	515	PDI_ASPNG	Q12730 aspergillus
8	163.5	3.9	504	PDA3_MOUSE	P11598 rattus norv
9	162	3.9	504	PDA3_MOUSE	P17967 saccharomyc
10	154.5	3.7	522	PDI_YEAST	P12865 trypanosoma
11	147	3.5	497	BS2_TRYBR	Q10057 schizosacch
12	146.5	3.5	492	PDI1_SCHPO	P36659 rattus norv
13	142	3.4	643	PDA4_RAT	P13667 homo sapien
14	138.5	3.3	645	PDA4_HUMAN	P38658 schistosoma
15	138	3.3	484	ER60_SCHMA	P38920 saccharomyc
16	137	3.3	769	MEH1_YEAST	P38194 saccharomyc
17	137	3.3	2493	YEA4_YEAST	P21195 oryctolagus
18	136	3.3	509	PDI_RABIT	P04785 rattus norv
19	135.5	3.3	509	PDI_RAT	P07237 homo sapien
20	135	3.3	508	PDI_HUMAN	P03102 galinus galli
21	135	3.3	510	PDI_BOVIN	P09103 mus musculu
22	134.5	3.2	517	BUG1_YEAST	O9x101 arabidopsis
23	134.5	3.2	638	PDA4_MOUSE	
24	133.5	3.2	361	PDA6_MOUSE	
25	132	3.2	2869	PDA6_AKATH	
26	131.5	3.2	364	RBP1_PLAVB	
27	131	3.2	1088	PDA6_MEDSA	
28	127.5	3.1	1905	RRPO_ROT51	
29	127	3.1	2108	Y659_PASMU	
30	126	3.0	493	NOFL_YEAST	
31	125	3.0	930	PDI_CHICK	
32	124.5	3.0	509	DPO1_HAETN	
33	123.5	3.0	501	PDI_MOUSE	
				PDI1_ARATH	

ALIGNMENTS

RESULT 1	ID	PDI_HUMAN	STANDARD:	PRT:	505 AA.
AC	P55059	PDI_HUMAN			
DT	01-OCT-1996	(Rel. 34, Created)			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	01-OCT-1996	(Rel. 34, Last annotation update)			
DE	Protein disulfide isomerase precursor (Ec 5.3.4.1) (PDI).				
OS	Humicola, Fungi; Ascomycota; mitosporic Ascomycota; Humicola.				
OX	NCBI_TaxID=34413;				
RM	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-KASI;				
RX	MEDLINE=94369094; PubMed=7765273;				
RA	Kajino T., Sarai K., Imada T., Idekoba C., Asami O., Yamada Y.,				
RA	Hirai M., Ueda S.;				
RT	"Molecular cloning of a fungal cDNA encoding protein disulfide				
RT	isomerase.";				
RT	Biosci. Biotechnol. Biochem. 58:1424-1429(1994).				
CC	-I- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING				
CC	DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION. PROLYL				
CC	HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).				
CC	-I- CATALYTIC ACTIVITY: Rearrangement of both intrachain and				
CC	interchain disulfide bonds in proteins to form the native				
CC	structures.				
CC	-I- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).				
CC	-I- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL: S74296; AAC60578.1; -				
DR	HSSP: P07237; IMEX				
DR	InterPro: IPR000886; ER_target.				
DR	InterPro: IPR000063; ThioRed.				
DR	Pfam: PF00085; ThioRed. 2.				
DR	PRINTS: PR00421; THIOREDOXIN.				
DR	TIGRFAMS: TIGR01126; pdi_dom; 2.				
DR	TIGRFAMS: TIGR01130; ER_pdi_fam; 1.				
DR	PROSITE: PS00014; ER_TARGET; 1.				
DR	PROSITE: PS00194; THIOREDOXIN; 2.				
KW	Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.				
FT	SIGNAL	1	20	POTENTIAL.	
FT	CHAIN	21	505	PROTEIN DISULFIDE ISOMERASE.	
FT	DISULFID	50	53	REDOX-ACTIVE (BY SIMILARITY).	
FT	DISULFID	385	388	REDOX-ACTIVE (BY SIMILARITY).	
FT	SITE	502	505	PREVENT SECRETION FROM ER (POTENTIAL).	
SO	SEQUENCE	505 AA;	55114 MW;	2767CODA16B5F95 CRC64;	

Query Match 4.5%; Score 185; DB 1; Length 505;

Matches 99; Conservative 96; Mismatches 180; Indels 122; Gaps 25;

QY 373 ELTVLEFETRNATV-----MASDIYL--FYAGMVAASMAFLOSIVDAVKL-KGST 423
 Db 20 EAENVVEGVLVAIVDNFKQLIADNEFVLYEYAPWCCHCKALADYAKAAQOLEKESP 79
 QY 424 MLITRNADMSDVCTKONTEFFPIIKMKYKGENPVSAVAGTDLKFIOLNRIISYPV 483
 Db 80 IKLAVDVTVEGELAEQYAVNRGYPLTKFFRSQ-SPEVYSGGQADIIAMVTKKTPAK 138
 QY 484 NITSIOAEVEYISGELYDYLLYSSVVLGFSPTMKTAKEDEFSAGNYLKGVIITGIYS 543
 Db 139 DLTSVADAEOFL-----KD-----NEIATIGFFKDESEBAKTFKYANALDSFVF-GVSS 188
 QY 544 EEDVLLSTKYAASLPALLLNRHTEGKTESIPLASTHODIYOITDALLEFPPEIYEN 603
 Db 189 NADVY--AKTEAKONGVYLFKFPDCK-----KSVFE--GELINEIN 224
 QY 604 LPSEYFRLOK-PULLIFSDGTVPQY-----KKAILTVKOKYLDSETPCWLNKTPVGRG 658
 Db 225 LKFAQVQSLPLIVDFNHSASKIRGSIKSHLFFVSE-----GG 266
 QY 659 ILRAFDPLPLPL-----LVLVNLSGGQ-----VFAP----- 687
 Db 267 HIEKVDPLKEIAKKYRDDILFVTISSEEDHTRIFEEFGMKKEEVPITRIKLEEDMAK 326
 QY 688 --PSQAIIEENLVLMKK-LEAGLENNHTI--LPAQEW-KPP--LPAVDPLSM-IDAA 737
 Db 327 YKPEDEDISAETIEAFLEFKFLDGKIKHLSOELP-EDMDKNPVYLVSSNESVALDLS 385
 QY 738 TSQGRTRKVP--KCMKETDVOENDEQHEDEKSAV-----RKPEIETLRKIHNNRSM 787
 Db 386 KSVLEEFAPMGCHCKQIAPYDQIAEKYKQMEDIVIAKMOSTANLESIKISSPTIKY 445
 QY 788 FKEAK-----SFRROKEL 801
 Db 446 FRKEDNKVIDENLDRTL 462

RESULT 3
 PDA3_BOVIN STANDARD; PRT; 505 AA.
 ID PDA3_BOVIN STANDARD; PRT; 505 AA.
 AC P3657;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein disulfide isomerase A3 precursor (Ec 5.3.4.1) (Disulfide isomerase ER-60) (ERP60) (58 kDa microsomal protein) (P58) (ERP57).
 GN PDI3 OR GRP58.
 OS Bos taurus (Bovine).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN NCBI_TaxID=9913;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96096758; PubMed=8529662;
 RA Hirano N., Shibasaki F., Sakai R., Tanaka T., Nishida J., Yasaki Y., Takenawa T., Hirai H.,
 RT "Molecular cloning of the human glucose-regulated protein ERp57/GRP58, a thiol-dependent reductase. Identification of its secretory form and inducible expression by the oncogenic transformation.".
 RT Eur. J. Biochem. 234:336-342(1995).
 CC -I- CATALYTIC ACTIVITY: Rearrangement of both intrachain and interchain disulfide bonds in proteins to form the native structures.
 CC -I- SUBCELLULAR LOCATION: Endoplasmic reticulum. lumen (By similarity).
 CC -I- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -I- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE TYPE I (PHOSPHOLIPASE C-ALPHA).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D16235; BAA03760.1; -
 DR HSRP: P07237; IMEX.
 DR InterPro: IPR000063; Thioried.
 DR Pfam: PF00085; Thioried; 2.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 KM Redox-active center; Isomerase; Endoplasmic reticulum. Repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 505 PROTEIN DISULFIDE ISOMERASE A3.
 FT DISULFID 57 60 REDOX-ACTIVE (BY SIMILARITY).
 FT DISULFID 406 409 REDOX-ACTIVE (BY SIMILARITY).
 FT SITE 502 505 PREVENT SECRETION FROM ER
 FT (BY SIMILARITY).
 SQ SEQUENCE 505 AA; 56929 MW; A85911748DCT0A23 CRC64;

Query Match 4.2%; Score 175.5; DB 1; Length 505;
 Best Local Similarity 23.3%; Pred. No. 0.00037;
 Matches 66; Conservative 56; Mismatches 128; Indels 33; Gaps 9;

QY 376 VELTEETRNATYMASDS---IVLYAGQAVSMAFLOSIVDAVKLKTSTMLTRING 431
 Db 28 LELTDDNFESRITLDGSSGLMVEFPAPWCCHCKALADYEAARLKG--IVPLAKVDC 85
 QY 432 ADMSVCTKQNTTEPPIIKMKYKGENPVSAVAGTDLKFIOLNRIISYPNITSORA 491
 Db 86 TANITCNKYGVSYPITLTFRDGEGSAYDPRADGIVSHLKKQACPAVPLKSEEF 145
 QY 492 EBYLSGELYKDLILYSSVVLGFSPTMKTAKEDESEAGNYLKGVIITGIYSEEDVLLS 551
 Db 146 EKFIISK-----DASVGFKDLFSEASESEFLKASNSNRD--NRFPHNTVESLY 193
 QY 552 TKYASLPALLAR--HTEGKTESIPLASTHAQ---DIVQIITDALLEFPPEIYENLP 605
 Db 194 NKYDDDEGICITLFRSHLNKFEKEDKVAYTEQKMTSGKIKRIFOENIFGICPHMEDND 253
 QY 606 SVFRLO-KPLILFSD-----GTVPQYKALITLVKQYILDS 642
 Db 254 ---LLOGKDLTAYVDYDEKNAGKSNINRNRYMAAKFFLDA 293

RESULT 4
 PDI_ASPOR STANDARD; PRT; 515 AA.
 ID PDI_ASPOR STANDARD; PRT; 515 AA.
 AC 000248;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Protein disulfide isomerase precursor (Ec 5.3.4.1) (PDI).
 GN PDI1.
 OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 CC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 CC NCBI_TaxID=5062;
 RN NCBI_TaxID=5062;
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIB 40;
 RA Lee B., Yamada O., Kitamoto K., Takahashi K.;
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
 CC -I- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 CC DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION. PROLYL
 CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
 CC -I- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 CC interchain disulfide bonds in proteins to form the native
 CC structures.

Query Match	4.2%;	Score 172.5;	DB 1;	Length 515;
Best Local Similarity	19.9%;	Pred. No. 0.00061;		
Matches 105;	Conservative 107;	Mismatches 185;	Indels 131;	Gaps 28

Db 114 DKGPNVDGGRDAGIVENWESRYDPNKKPPEVEVTLTTEENDDPISNNELVLEFYA 173
 QY 399 GMAQVSMAFLOSIDYDAVAKIKGT-STMLTFRINCADMSDVCQKQVTEPPIIKMKKKKEN 457
 Db 174 PWCIGCKLAPLEYEKAQKAKQSKVKLGKVDATIEKDLGKGVSGIPMKIIRNGRR 233
 QY 458 PVSVAAGMLGTDLKFLIOLNRIISYPVNTISIOAEYLSGELIKDLIIYSVSVGLFSP 517
 Db 224 -FDVNGPREAAGIIKYMTPDQSKPAKKLPLKLDVERFMS-----KD-----DVTIIGFAT 283
 QY 518 TMTAKKEPSEAGNLYKGVITGIISSE-----DVLISLTKIASLPDLLAR 565
 Db 284 EDSTAFEAFSDAEMLRREFTMGHTSDPAFKKWDAPKNDIIF-----YPSLFSK 336
 QY 566 HTEGKIESIPLASTFAODIYOITPDLALPEPTEIVENLPSYFRLOKPLLIIFSGTVPN 625
 Db 337 F-EPKSRITNNKAAATSEDLAFREHSAPIYGMKTKNAATY-TRKPLVYVYVYNDASV 394
 QY 626 QYKKA-----ILTVKQKYLDSFTPCWMLNKTPVGRGILRAYFDPLPLLVNL 678
 Db 395 QYREGSEYWRSKVLINAKQYKDKYK----- 420
 QY 679 HSGGVAFEPDQATIEENLVMLKLL---EAGLENHTITLPAQEMKPPPLPYDLSMID 735
 Db 421 -----FAVADDEEFKAE-----LEELGLDGSGLHNHYVFGYDGKKYPMNDEDFEID 469
 QY 736 ----AATSGRGTRKVRKCMKERTVOENDK 760
 Db 470 ENLEAFMKQISSGKAKAHVKSAPAKDK 498

RESULT 6

PD03_HUMAN STANDARD: PRT: 505 AA.
 AC P30101; Q14255; Q13453; Q9UMU7;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein disulfide isomerase A3 precursor (BC 5.3.4.1) (Disulfide isomerase ER-60) (ERP60) (58 kDa microsomal protein) (p58) (Erp57)
 DE (58 kDa glucose regulated protein).
 GN PDIA3 OR GRP58 OR ERP60.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95032122; PubMed=7945384;
 RA Hirano N., Shibasaki F., Katoh H., Sakai R., Tanaka T., Nishida J., Yasaki Y., Takenawa T., Hirai H.;
 RT "Molecular cloning and characterization of a cDNA for bovine phospholipase C-alpha: proposal of redesignation of phospholipase C-alpha.";
 RL Biochem. Biophys. Res. Commun. 204:375-382(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Charneck-Jones D.;
 RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kito M., Urae R.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=96063616; PubMed=7487104;
 RA Bourd M., Demady D., Martin J.L., Jabbour S.K., Martin B.M., George J.W., Pohl U.R.;
 RT "cDNA cloning and baculovirus expression of the human liver endoplasmic reticulum p58: characterization as a protein disulfide isomerase isoform, but not as a protease or a carnitine

RT acyltransferase.";
 RL Arch. Biochem. Biophys. 323:397-403(1995).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97349107; PubMed=9205111;
 RA Kojunen P., Horelli-Kuitunen N., HeLaakoski T., Karyonen P., Jaakkola M., Palotie A., Kivirikko K.I.;
 RT "Structures of the human gene for the protein disulfide isomerase-related polypeptide ERp60 and a processed gene and assignment of these genes to 15q15 and 1q21.";
 RL Genomics 42:397-404(1997).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 25-33.
 RC TISSUE=Liver;
 RX MEDLINE=93162045; PubMed=1286669;
 RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F., Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargias R., Appel R.D., Hughes G.J.;
 RT "Human liver protein map: a reference database established by microsequencing and gel comparison.";
 RL Electrophoresis 13:992-1001(1992).
 RN [8]
 RP SEQUENCE OF 26-42.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=97295304; PubMed=9150946;
 RA Rasmussen R.K., Ji H., Edes J.S., Moritz R.L., Reid G.E., Simpson R.J., Dorow D.S.;
 RT "Two-dimensional electrophoretic analysis of human breast carcinoma proteins: mapping of proteins that bind to the SH3 domain of mixed lineage kinase MLK2.";
 RL Electrophoresis 18:588-598(1997).
 RN [9]
 RP SEQUENCE OF 25-54; 62-75 AND 95-104.
 RC TISSUE-Colon carcinoma;
 RX MEDLINE=97295306; PubMed=9150948;
 RA Ji H., Reid G.E., Moritz R.L., Edes J.S., Burgess A.W., Simpson R.J.;
 RT "A two-dimensional gel database of human colon carcinoma proteins.";
 RL Electrophoresis 18:605-613(1997).
 RN [10]
 RP SEQUENCE OF 95-104 AND 472-479.
 RC TISSUE-Keratinocytes;
 RX MEDLINE=93162043; PubMed=1286667;
 RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E., Vandekerckhove J.;
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes.";
 RL Electrophoresis 13:960-969(1992).
 RN [11]
 RP MASS SPECTROMETRY.
 RC TISSUE=Breast cancer;
 RX MEDLINE=21829512; PubMed=11840567;
 RA Harris R.A., Yang A., Stein R.C., Lucy K., Brusten L., Herath A., Parekh R., Waterfield M.D., O'Hare M.J., Neville M.A., Page M.J., Zvelebil M.J.;
 RT "Cluster analysis of an extensive human breast cancer cell line protein expression map database.";
 RL Proteomics 2:212-223(2002).
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and interchain disulfide bonds in proteins to form the native structures.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
 CC -1- MASS SPECTROMETRY: MW=54265.22; METHOD=MALDI.
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE TYPE I (PHOSPHOLIPASE C-ALPHA).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D16234; BAA03759.1; -
 DR EMBL: 249835; CAA89996.1; -
 DR EMBL: D83485; BAA11928.1; -
 DR EMBL: U42068; AAC50331.1; -
 DR EMBL: U75885; AAC51518.1; -
 DR EMBL: U75875; AAC51518.1; JOINED.
 DR EMBL: U75876; AAC51518.1; JOINED.
 DR EMBL: U75877; AAC51518.1; JOINED.
 DR EMBL: U75878; AAC51518.1; JOINED.
 DR EMBL: U75879; AAC51518.1; JOINED.
 DR EMBL: U75880; AAC51518.1; JOINED.
 DR EMBL: U75881; AAC51518.1; JOINED.
 DR EMBL: U75882; AAC51518.1; JOINED.
 DR EMBL: U75883; AAC51518.1; JOINED.
 DR EMBL: U75884; AAC51518.1; JOINED.
 DR EMBL: BC014433; AAH14433.1; -
 DR HSSP: P07237; IMEX.
 DR MEROPS: C17.001; -
 DR SWISS-2DPAGE: P30101; HUMAN.
 DR Aairhus/Ghent-2DPAGE; 5410; IEF.
 DR PHCI-2DPAGE: P30101; -
 DR Slena-2DPAGE: P30101; -
 DR Genew: HGNC:4606; GRP58.
 DR MIM: 602046; -
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 2.
 DR PRINTS: PRO0421; THIOREDOXIN.
 DR TIGRERAMS: TIGR01126; pdi_dom; 2.
 DR TIGRERAMS: TIGR01130; ER_PDI_fam; 1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 DR Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 KW KW
 FT SIGNAL 1 24
 FT CHAIN 25 505
 FT DISULFID 57 60
 FT DISULFID 406 409
 FT SITE 502 505
 FT
 FT CONFLICT 19 19
 FT CONFLICT 22 22
 FT CONFLICT 217 217
 FT CONFLICT 225 225
 FT CONFLICT 238 238
 FT CONFLICT 272 272
 FT CONFLICT 355 355
 FT CONFLICT 358 358
 FT CONFLICT 368 368
 SQ SEQUENCE 505 AA; 56782 MW; 529E5B692D0D7E9 CRC64;

Query Match 4.1%; Score 169.5; DB 1; Length 505;
 Best Local Similarity 22.6%; Pred. No. 0.00093;
 Matches 64; Conservative 59; Mismatches 127; Indels 33; Gaps 10;

QY 376 VELTEETFNATVWASDS---YLFYAGQWQAVMAFLQSYIDVAVKLGKSTMLTRINC 431
 DB 28 LELTDNFSRISDTSAGLIMLVEFFAPMGCHCKRLAPELEAATRLKG--IYPLAKVDC 85
 QY 432 AWMSDVCTKQNTVEPPIIKMYKKGKGNPVSYAGMLGTRKDLKFTQLNRISYPVNTSIQEA 491
 DB 86 TANTNTCNKYGVSGLYPLKIFRDGEAGAYDGPRTADIVS--HLKKQAGPASPVLRTTEE 143
 QY 492 EETLSEELKDLILYSSVSLGLFSPMTAKADESEAGYLYLGVIYITGTYSEEDVLLS 551
 DB 144 E-----FRKFLSDKDAIVGFFDSSFSEAHSEFLKAASNLRD---NRFATINVESTIV 193
 QY 552 TRYAAALPALLLAR--HTGEKIESIPLASTHAQ---DIYQITDALLLEMFPEITVENLP 605
 DB 194 NEEDDNGEGILLFRPSHLTKKFEEDKTYVAITEQMTSGKIKKFKIQENIFGICPHATEDNKD 253

QY 606 SYPRLO-KPLLIFSD-----GTVPQYKKAITLYKQKYLDS 642
 DB 254 ---LIQKDLLILYVDYDEKNAKKSNNYMRNRYMVAKKRFLDA 293

RESULT 7
 PDI ASPNG
 ID PDI ASPNG STANDARD; PRT; 515 AA.
 AC 012730;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Protein disulfide isomerase precursor (EC 5.3.4.1) (PDI).
 GN PDI OR PDI.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 120.49 / N400;
 RX MEDLINE=97174110; PubMed=9021130;
 RA Ngiam C., Jeenes D.J., Archer D.B.;
 RT Isolation and characterisation of a gene encoding protein disulphide
 RT isomerase, pdiA, from Aspergillus niger.
 RL Curr. Genet. 31:133-138(1997).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 3;
 RA Walplich S.;
 RL Submitted (Jul-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 CC DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
 CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 CC interchain disulfide bonds in proteins to form the native
 CC structures.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X89797; CAA67332.1; -
 DR EMBL: X89449; CAA61619.1; -
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 3.
 DR PRINTS: PRO0421; THIOREDOXIN.
 DR TIGRERAMS: TIGR01126; pdi_dom; 2.
 DR TIGRERAMS: TIGR01130; ER_PDI_fam; 1.
 DR PROSITE: PS00014; ER_TARGET; 1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 KW KW
 FT SIGNAL 1 20
 FT CHAIN 21 515
 FT DISULFID 54 57
 FT DISULFID 389 392
 FT SITE 512 515
 SQ SEQUENCE 515 AA; 56291 MW; 2B005B8786400AD9 CRC64;

Query Match 4.1%; Score 169.5; DB 1; Length 515;
 Best Local Similarity 20.6%; Pred. No. 0.00095;
 Matches 111; Conservative 89; Mismatches 193; Indels 145; Gaps 27;

```

Db 44 LVLAFFPWCCHKALAP-KY-----EBAATELAKAKNIPLYVDCSTAE----- 87
QY 75 CGKEDLMKAVIFKGNIIILREF-----PDTLFDVAIVAHVIFALL--FSEKYYTN 125
Db 88 ----DLRSQVEVEYPTLKIFRGVDSKPYOGAROTESIYSYMKOSIPAVSVNE-EN 141
QY 126 LEDLONINNALKGKANIIFSYVRAIGIEPHRAVMEAGFYGGTYQFVLTTEALLESIGS 185
Db 142 LEEITMKI-----VIGIYPSDDOETQAFEKYAESQRNRYLFAATDDAAIAIKSEV 195
QY 186 EDVEYAHLY--FPHCKLVLD--LTQOCRTIMEOPITLNLHLFKTKMAKPLLEVAEDP 241
Db 196 EOPSTIV-LYKDEDEKAVYDGEIEDEA-----HISWKSASTPLVGEIG--P 239
QY 242 QOVSVTHQLGLPLVFIYSQATYFADRTAEW-----VAMPLLG-KACV 285
Db 240 ETVSG-YIGAGVPLAYIFA-ETKEREKRYTEDFKPAQKHKAINIATIDAKMFAHAGN 297
QY 286 LLLRDSLEVNIPD-----ANVYKRAEE--GVPEFLVHDVDLIIHVENNMHIEEQ 339
Db 298 LNLDSQKPPARAIDPPANNAKPYDQAKELNADVEKFT--QDVLGKVEPSISSEVP 354
QY 340 EDEDNDMEGPDIDVODDEVAETVFRDRKRLPLELTVLEETFEFNATWASDSIYL--FY 397
Db 355 ESQ-----EGP-----VTV--VVAHSYKDIYDNDKDVLEFY 385
QY 398 AGQAVSAFPIQSYIDVAVKLGSTMTLITRINCADMSDVCYK-----ONV 443
Db 386 APWCHCKALAPKYDELALY-----ADHPDLAAKAVIYAKIDATANDVPPPI 432
QY 444 TEFLPKMYKKG--ENPVSYAGMLGTDKDLKPIOLNRSYPNITSIDEAEFYLSGEL 499
Db 433 TGFPRLRLYPGAKADSPLEYSGSRTEVDLANFVKENG-KHNDALNVASEEQEGDV 489

```

```

RL Biochem. Biophys. Res. Commun. 178:679-685(1991).
RN [4]
RP SEQUENCE OF 26-43.
RC TISSUE=Brain, and pituitary;
RX MEDLINE=90208308; Pubmed=2181662;
RT Mobs C.V., Flak G., Pfaff D.W.;
RT "HIP-70": a protein induced by estrogen in the brain and LH-RH in the
RN pituitary";
RN Science 247:1477-1479(1990).
RP [5]
RP SEQUENCE OF 25-54; 258-269; 285-310; 347-350; 412-419 AND 434-463.
RC TISSUE=Liver;
RX MEDLINE=92041865; Pubmed=1657921;
RA Srivastava S.P., Chen N.O., Liu Y.X., Holtzman J.L.;
RT "Purification and characterization of a new isozyme of thiol:protein-
RT disulfide oxidoreductase from rat hepatic microsomes. Relationship of
RT this isozyme to cytosolic phosphatidylinositol-specific phospholipase
RT C form 1A.";
RL J. Biol. Chem. 266:20337-20344(1991).
RN [6]
RP SEQUENCE OF 26-34; 174-193; 433-446 AND 448-458.
RX MEDLINE=92340568; Pubmed=1321829;
RA Urade R., Nasu M., Moriyama T., Wada K., Kito M.;
RT "Protein degradation by the phospholipase-specific phospholipase
RT C-alpha family from rat liver endoplasmic reticulum.";
RL J. Biol. Chem. 267:15152-15159(1992).
RN [7]
RP INHIBITION BY PHOSPHOLIPIDS.
RX MEDLINE=93050170; Pubmed=1330685;
RA Urade R., Kito M.;
RT "Inhibition by acidic phospholipids of protein degradation by ER-60
RT F6B5 lect. 312:83-86(1992).
RL [8]
CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
CC interchain disulfide bonds in proteins to form the native
CC structures.
CC -1- ENZYME REGULATION: SEEMS TO BE INHIBITED BY ACIDIC PHOSPHOLIPIDS.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A PHOSPHATIDYL-
CC INOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE TYPE I (PHOSPHOLIPASE
CC C-ALPHA) THEN WAS THOUGHT (REF.6 AND REF.7) TO BE A THIOL
CC PROTEASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X12355; CAA30916.1; -.
DR EMBL: D63378; BAA09695.1; -.
DR PIR: A28807; A28807.
DR PIR: A40095; A40095.
DR HSRP: P07237; IMEX.
DR MEROPS: C17.001; -.
DR InterPro: IPR000063; ThioRed.
DR Pfam: PF00085; ThioRedoxin.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRFSMS: TIGR01126; pdl_dom; 2.
DR TIGRFSMS: TIGR01130; ER_PDI_fam; 1.
DR PROSITE: PS00194; THIOREDOXIN; 2.
KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 24
FT CHAIN 25 505
FT DISULFD 57 60
FT DISULFD 406 409
FT SITE 502 505
FT CONFLICT 1 13
FT 1).

```

```

FT CONFLICT 98 98 S -> T (IN REF. 1)..
FT CONFLICT 232 240 IKK1QEST -> SRSLFRKA (IN REF. 1).
FT CONFLICT 476 476 F -> L (IN REF. 2).
SQ SEQUENCE 505 AA; 56623 MW; EAC7FC0BD4F1471 CRC64;

Query Match
Best Local Similarity 20.6%; Pred. No. 0.0023; DB 1; Length 505;
Matches 61; Conservative 56; Mismatches 120; Indels 59; Gaps 9;

OY 376 VELTEEFNATVWASDS---IVLFYAGQVMAFLQSYIDVAVKIKGTSTMLTRINC 431
DB 28 LETIDENFSRVSDTGSAGLMLVEFPAPCGHCKRLAPEEAATRLKG--IYPLAKVDC 85
OY 432 ADMSDVCTQNVTEFFPIIKYKKGENVSYAGMLGTRDKLFLQLNRISPVNITSIOEA 491
DB 86 TANTNCTNKNYVSGVPTLKFIRGEEAGVADGPRADGIVS--HLKKQAPASVPLRTED 143
OY 492 EEYLSGELKDLIYSSVYLGLFSPMTKAKEDFSEAGNYLKGV-----TTGIYSEE 545
DB 144 E-----FKKFIISDKDASVGFRRDLSFGSHSEFLKAASNLDRNFRANTVSELYKEY 196
OY 546 D-----VLLSTKYAASLPALLARHTEGKIESIPLASTHADIQIITDALLEM 595
DB 197 DNGEGITIFRPLHANKFEDKIVAYTEKKMTSGKIK-----KFIQESIFGL 243
OY 596 FEPIITVEN-----LPSYFRLQKPLILFSDGTVNPOYKKAILTLYKQKILDS 642
DB 244 CPHMTEDNKDLIOGKDLTAAYDVD-----YEKNKGSNMYNRNRYMVAKKFLDA 293

RESULT 9
PDA3_MOUSE STANDARD: PRT; 504 AA.
ID PDA3_MOUSE
AC P27773;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein disulfide isomerase A3 precursor (EC 5.3.4.1) (Disulfide
isomerase ER-60) (ERP60) (58 kDa microsomal protein) (p58) (ERP5).
GN PDIA3 OR GRP58 OR ERP60 OR ERP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91237072; PubMed=2033248;
RX Hempel W.M., DeFranco A.L.;
RT "Expression of phospholipase C isozymes by murine B lymphocytes.";
RL J. Immunol. 146:3713-3720(1991).
RN [2]
RP SEQUENCE OF 25-41.
RC TISSUE=Fibroblast;
RX MEDLINE=9500907; PubMed=7523108;
RA Merrick B.A., Paterson R.M., Michler L.L., He C., Selkirk J.K.;
RT "Separation and sequencing of familial and novel murine proteins
using preparative two-dimensional gel electrophoresis.";
RL Electrophoresis 15:735-745(1994).
RN [3]
RP SEQUENCE OF 25-41, AND PHOSPHORYLATION.
RC TISSUE=Fibroblast;
RA Merrick B.A., Michler L.L., Paterson R.M., He C., Selkirk J.K.;
RT "Identification of the two isoforms of phospholipase C-alpha from
dividing murine fibroblasts by protein microsequencing.";
RL Biochem. Arch. 9:335-340(1993).
CC -I- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
interchain disulfide bonds in proteins to form the native
structures.
CC -I- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
CC -I- PTM: PHOSPHORYLATED.
CC -I- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC -I- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PHOSPHATIDYLINOSITOL-4,5-
BISPHOSPHATE PHOSPHODIESTERASE TYPE I (PHOSPHOLIPASE C-ALPHA).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M73329; AAA39944.1; -.
DR HSSP: P07237; IMEK.
DR SWISS-2DPAGE: P27773; MOUSE.
DR MGD: MGT:95834; GTP58.
DR InterPro: IPR000063; ThioRed.
DR Pfam: PF00085; thioRed; 2.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRPFAMs: TIGR01126; pdi.dom; 2.
DR TIGRPFAMs: TIGR01130; ER_pdi.fam; 1.
DR PROSITE: PS00194; THIOREDOXIN; 2.
KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal;
KW Phosphorylation.
FT SIGNAL 1 24
FT CHAIN 25 504 PROTEIN DISULFIDE ISOMERASE A3.
FT DISULFID 57 60 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 405 408 REDOX-ACTIVE (BY SIMILARITY).
FT SITE 501 504 PREVENT SECRETION FROM ER
FT SITE (BY SIMILARITY).
SQ SEQUENCE 504 AA; 56621 MW; 59B3B04755CEB7B4 CRC64;

Query Match
Best Local Similarity 21.8%; Pred. No. 0.0029;
Matches 63; Conservative 55; Mismatches 125; Indels 46; Gaps 10;

OY 376 VELTEEFNATVWASDS---IVLFYAGQVMAFLQSYIDVAVKIKGTSTMLTRINC 431
DB 28 LETIDENFSRVSDTGSAGLMLVEFPAPCGHCKRLAPEEAATRLKG--IYPLAKVDC 84
OY 432 ADMSDVCTQNVTEFFPIIKYKKGENVSYAGMLGTRDKLFLQLNRISPVNITSIOEA 491
DB 85 TANTNCTNKNYVSGVPTLKFIRGEEAGVADGPRADGIVS--HLKKQAPASVPLRTED 142
OY 492 EEYLSGELKDLIYSSVYLGLFSPMTKAKEDFSEAGNYLKG---YYTGIYSEEDVL 548
DB 143 E-----FKKFIISDKDASVGFRRDLSFGSHSEFLKAASNLDRNFRANTVSELYKEY 190
OY 549 LSTKYAASLPALLAR--HTEGKIESIPLASTH---ADIQIITDALLEMFEPIIVE 602
DB 191 -LVKEYDDNGEGITIFRPLHANKFEDKIVAYTEKKMTSAKIKFIQDSIFGLCPHMTED 249
OY 603 N-----LPSYFRLQKPLILFSDGTVNPOYKKAILTLYKQKILDS 642
DB 250 NKDLIOGKDLTAAYDVD-----YEKNKGSNMYNRNRYMVAKKFLDA 292

RESULT 10
PDI_YEAST STANDARD: PRT; 522 AA.
ID PDI_YEAST
AC P17967;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Thioredoxin-
related glycoprotein 1).
GN PDI1 OR MFP1 OR TRG1 OR YC1043C OR YC143C OR YC1313.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=X2180-1A, and TMS;
RX MEDLINE=92105067; PubMed=1761527;
RA Tachikawa H., Miura T., Katakura Y., Mizunaga T.;

```

RT "Molecular structure of a yeast gene, PDI1, encoding protein
 RT disulfide isomerase that is essential for cell growth.";
 RL J. Biochem. 110:306-313(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91289690; PubMed-2063627;
 RA Scherens B., Dubois E., Messenguy F.;
 RT "Determination of the yeast YOL131 gene localized on
 RT chromosome III. Homology with the protein disulfide isomerase (PDI
 RT gene product) of other organisms.";
 RL Yeast 7:185-193(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91239586; PubMed-1840696;
 RA Imanita M., Miura T., Tachikawa H., Kaplan H.A., Lennarz W.J.,
 RA Mizunaga T.;
 RT "Glycosylation site binding protein and protein disulfide isomerase
 RT are identical and essential for cell viability in yeast.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4453-4457(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92105117; PubMed-1761554;
 RA Gentner R., Brauer C., Janetzky B., Foerster H.H., Ehbrecht I.M.,
 RA Lehle L., Kuentzel H.;
 RT "The Saccharomyces cerevisiae TRG1 gene is essential for growth and
 RT encodes a luminal endoplasmic reticulum glycoprotein involved in the
 RT maturation of vacuolar carboxypeptidase.";
 RL J. Biol. Chem. 266:24557-24563(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92397595; PubMed-1523890;
 RA Scherens B., Messenguy F., Gigot D., Dubois E.;
 RT "The complete sequence of a 9,543 bp segment on the left arm of
 RT chromosome III reveals five open reading frames including glucokinase
 RT and the protein disulfide isomerase.";
 RL Yeast 8:577-586(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92104510; PubMed-1761235;
 RA Fargnhar R., Honey N., Murtant S.J., Bossier P., Schultz L.,
 RA Montgomerie D., Ellis R.W., Freedman R.B., Rulle M.F.;
 RT "Protein disulfide isomerase is essential for viability in
 RT Saccharomyces cerevisiae.";
 RL Gene 108:81-89(1991).
 CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 CC DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
 CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER.
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 CC interchain disulfide bonds in proteins to form the native
 CC structures.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (potential).
 CC -1- PM: The N-terminus is blocked.
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52313; CAA36550.1; -
 DR EMBL: X57712; CAA40883.1; -
 DR EMBL: M62815; AAA34848.1; -
 DR EMBL: M76982; AAA35169.1; -
 DR EMBL: X59720; CAA42373.1; -
 DR EMBL: X54535; CAA38402.1; -
 DR EMBL: D00842; BAA00723.1; -
 DR PIR: JX0182; ISRYSS.
 DR PIR: S19372; S19372.
 DR PIR: S15050; S15050.
 DR PIR: S25349; S25349.

DR PIR: A41713; A41713.
 DR PIR: JS0634; JS0634.
 DR HSSP: P07237; IMEK.
 DR SGD: S0000548; PDI1.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 2.
 DR PRINTS: PR00421; THIAREDOXIN.
 DR TIGRFAMs: TIGR01126; pdi_dom; 2.
 DR TIGRFAMs: TIGR01130; ER_pdi_fam; 1.
 DR PROSITE: PS00194; ER_TARGET; 1.
 DR PROSITE: PS00194; THIAREDOXIN; 2.
 KM Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 522
 FT DISULFID 61 64
 FT DISULFID 406 409
 FT CARBOHYD 82 82
 FT CARBOHYD 117 117
 FT CARBOHYD 155 155
 FT CARBOHYD 174 174
 FT CARBOHYD 425 425
 FT SITE 519 522
 FT CONFLICT 33 52
 FT CONFLICT 83 83
 FT CONFLICT 114 114
 FT CONFLICT 143 143
 FT CONFLICT 146 146
 FT CONFLICT 168 168
 FT CONFLICT 197 197
 FT CONFLICT 215 215
 FT CONFLICT 226 226
 FT CONFLICT 333 333
 FT CONFLICT 351 351
 FT CONFLICT 455 455
 FT CONFLICT 458 458
 FT CONFLICT 505 505
 SQ SEQUENCE 522 AA; 58227 MW; 690CF3E05D7F74C94 CnC64;
 Query Match 3.7%; Score 154.5; DB 1; Length 522;
 Best Local Similarity 20.8%; Pred. No. 0.0094;
 Matches 66; Conservative 67; Mismatches 126; Indels 58; Gaps 15;
 QY 358 VAETFERDRKRRLPLE-LTVELTEETFNATWASQSYL-FAGQQAASMAFLQSYDVA 415
 DB 16 LASSVEFAQOEAAYAPEDSAVAVKATDSFNEYIOSHDLVAEFPAFCGCKKNAPETVRAA 75
 QY 416 VLKGTSTMLTRINCADMSDVCTKQNTPEPIIMKRGKE--NPVSAGMLGTDLKLF 473
 DB 76 ETLV-EKNITLQAIDCTENODLCMEHNIIPGFSLTIFPNSVDNNSIDYEGPTAAIYOF 134
 QY 474 IQLNRISTPVNTISIQEAEVYISGLYKDLIYSSVSLGLFSPMKAKEDFSAGNYL 533
 DB 135 --MIQSQPA-VAAYADDPAYLANETFTVPTVQSGKIDAEFNATF-----YSMAKHF 185
 QY 534 KGYVITGISSEEDVLLSTKYASL--PALLARHTTECKIESIPLASTHAQDIQIIDA 591
 DB 186 NDYDVSAMENADDDKSLIYPSANDEPVYVNGKAD-----LIDA 226
 QY 592 LLEMPPE-ITVENLPSYRLOKPLILFSDGTNPQYKAILTL-----VKQYLDSETP 645
 DB 227 --DVEFKWLOVEALPYFGEI-----DGSVFQAYVESGLPLGLYLNDEEELJEYKP 275
 QY 646 CWLNL--KNTPVGKIL 660
 DB 276 LFTELAKKN---RGIM 288
 RESULT 11
 BS2_TRYBB STANDARD; PRT; 497 AA.
 ID BS2_TRYBB
 AC P12865;

DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Bloodstream-specific protein 2 precursor.
 GN B52.
 OS Trypanosoma brucei brucei.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxId=5702;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90001196; PubMed=2551375;
 RA Hsu M.P., Muchic M.L., Boothroyd J.C.;
 RT "A developmentally regulated gene of trypanosomes encodes a homologue
 of rat protein-disulfide isomerase and phosphoinositol-phospholipase
 C.";
 RL Biochemistry 28:6440-6446(1989).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J02865; AAA30168.1; -
 DR PIR: A32820; A32820.
 DR HSSP: P07237; 1MER.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 3.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGR: TIGR01126; pdl_dom; 2.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 FT SIGNAL 1 14
 FT CHAIN 15 497
 FT DISULFID 48 51
 FT DISULFID 378 381
 FT CARBOHYD 30 30
 FT CARBOHYD 63 63
 FT CARBOHYD 85 85
 FT CARBOHYD 133 133
 FT CARBOHYD 154 154
 FT CARBOHYD 250 250
 FT CARBOHYD 278 278
 FT CARBOHYD 413 413
 FT CARBOHYD 465 465
 FT CARBOHYD 476 476
 FT CARBOHYD 482 482
 FT CARBOHYD 485 485
 FT CARBOHYD 488 488
 SQ SEQUENCE 497 AA; 55579 MW; 8A8EC4AD762B67EB CRC64;
 Query Match 3.5%; Score 147; DB 1; Length 497;
 Best Local Similarity 19.4%; Pred. No. 0.027;
 Matches 77; Conservative 76; Mismatches 158; Indels 86; Gaps 16;
 OY 375 TVELEETFNATYMSD-STVLFPYAG-----WQVSAVAFQSYIDVAVKLGTS 422
 DB 21 SLTLTRENENETIAKSEFLVRYVTCGYCQMLADEMEKAA---NETIDN----- 69
 OY 423 TMLTFRINCADSDVCTKQWTEFPPIIKMKKGNPVSAGMLGTDFLKLFIOLNRSYP 482
 DB 70 --LGMGVDCDSQPELANSSIRGYPITLLFRNGKEAHNGAGTKTDIIKIYKAN---VG 124
 OY 483 VNITSQDAEYVLSGLYDLIYSSVVLGFSPTMKAKEDFSAGNYLKGVTIGIY 542
 DB 125 PAVTPASNAEEVYRAKEEDHV-----CVGLTANNSTSTLTADAAQSF----- 170
 OY 543 SEEDVLLSTKYASLPALL-----LARHTGKTESIPLASTHMODIVQITTDALLEN 595

DB 171 -----VSLKFEAEAPKLPPEDEKPEITYVYRKGGKEYYDGMEEYKLTPELIQISRVAF 223
 OY 596 FPEITVENLPSPYRLOKPL--LTFSDGTVPQYKRAIITLVYKQYLDSEFTPCMLNKPT 653
 DB 224 GGEITPENQYQYVIRKPGWAMVKPNETASIELKES-LTEVGKKRSHVYLVNMSKH 282
 OY 654 PVCGRLIRAFVPLP---PLPLLVLVN-----LHSGGVFAFPASDQAIIENLVMLKL 705
 DB 283 PWRDR-----GVPEDAKYPALIAIHGMANYLHSTAEV-----VYRESLEKIFLEF 328
 OY 706 EAG-LENHTITLPAQEMKPLPAYDFLSMDATFSOR 741
 DB 329 AAGRVETIKSLPV-----PEVETVDKTTIVATMKK 361
 RESULT 12
 PDIL SCHPO
 ID PDIL SCHPO STANDARD; PRT; 492 AA.
 AC 010057;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative protein disulfide isomerase ClF5.02 precursor (EC 5.3.4.1).
 GN SPAC1F5.02.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX Schizosaccharomycetes.
 OX NCBI_TaxId=4896;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,
 RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Wellner-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leilaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROXYL
 HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 interchain disulfide bonds in proteins to form the native
 structure.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; Z68136; CA92230.1; -.
DR HSSP; P07237; IMK.
DR InterPro; IPR000886; ER-target.
DR InterPro; IPR00063; ThioRed.
DR Pfam; PF00085; thioRed; 2.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRPFAMS; TIGR01126; pdl_dom; 2.
DR TIGRPFAMS; TIGR01130; ER_pdl_fam; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00194; THIOREDOXIN; 2.
KW Hypothetical protein; Redox-active center; Isomerase;
KW Endoplasmic reticulum; Repeat; signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 492 PUTATIVE PROTEIN DISULFIDE ISOMERASE
FT CDS 51 54 CDS.02.
FT DISULFID 385 388 REDOX-ACTIVE (BY SIMILARITY).
FT CARBOHYD 161 161 REDOX-ACTIVE (BY SIMILARITY).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 489 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 492 AA; 54880 MW; C50B59236566667 CRC64;

Query Match 3.5%; Score 146.5; DB 1; Length 492;
Best Local Similarity 19.3%; Pred. No. 0.029;
Matches 103; Conservative 71; Mismatches 186; Indels 175; Gaps 20;

QY 15 VINCIFMPTVNSLPDELSPQKFTSLQGLLENEAVRPLDQGISVAKVNCVKR----- 69
DB 41 VLAKVFVAPWCCHCKALP-----EYESADELEKDGSLVEVDCETEGDGLCS 88
QY 70 -----EISRCCKEK-DLAKAVLFKGNL-LAREPDTLFDVAIAVAH 110
DB 89 EYSIRGYPVLNVFKNGKQISQYSGPRKHDAVLVKRKQDLPTVKDISKDTL----- 139
QY 111 VLFALLFESEVKYITNLEPLQNIENALKGNANIFSYVAIGIPERAVAEAGFYVGTYYQ 170
DB 140 -----ENFVEKADLAVVAFPKDKRLNDTY-----EVAEYWKDFV----- 176
QY 171 FVLTEITALLESIGSEDEVAVHALYFPHCKLVLDLTQOCRRPLMEOPLTTLNHLFEIKTK 230
DB 177 FAASDCKELAKSLGS--NPGIYAF-----TKDAQDSQKLVYTGDMWPASIADELTGVSS 229
QY 221 APLTEVAEDQOVSTVHLQGLPLVFI----- 258
DB 230 IPLDEL---NQMTFGKYOQSGPLGIIFYNSTESRDELYDFOPLAKKYQDTLRFAPLD 286
QY 259 -----VSGQATYEADRTAEVAVARLLGKAGVLLLRDSLEVNIPQDANVYKRAEGV 312
DB 287 AVRGCAVAKQANVESD-----WPAF-----VIANLSMKLVYPPR--TTELTAKMKTFV 333
QY 313 PVEFLVLDVLDLISHVENNMHIEIOEDENDEMGPPIDVQDDVAVETVFRKKRKLPL 372
DB 334 G-----DFVQSKLQPKIKSQPIPEQED----- 356
QY 373 ELTYELTEETENATVM--ASDSIVLFYAGQAVSAPLQSYIDAVVKLKGSTMLLTRIN 430
DB 357 --LVVLVADNPDIDYMDETKDYLVYEFVAPWCCHCKNLAPTYEKLAEEYSDSQSNVAVAKID 414
QY 431 CADMSDVCQKQNVTEPFIKKMYKKE--NPVSYAGMLGKDKLQKLTQUNRISYPV 483
DB 415 ATE-NDISV--SISGFTIMEFKANDKVPVAREGDRLELDSAFIDKHASEPI 466

RESULT 13
PDA4_RAT STANDARD: PRT; 643 AA.
AC P36659;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-Oct-2001 (Rel. 40, Last annotation update)

```

DE Protein disulfide isomerase A4 precursor (EC 5.3.4.1) (Protein Erp-72)

DE (Erp72) (Calcium-binding protein 2) (Cabp2).

GN PDIA4 OR CABP2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Liver;

RX MEDLINE=93238767; PubMed=8477750;

RA Van P.N., Rupp K., Lampen A., Soelling H.-D.;

RT "Cabp2 is a rat homolog of Erp72 with protein disulfide isomerase

activity";

RL Eur. J. Biochem. 213:789-795(1993).

CC -1 CATALYTIC ACTIVITY: Rearrangement of both intrachain and

interchain disulfide bonds in proteins to form the native

structures.

CC -1 SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.

CC -1 INDUCTION: UPON GLUCOSE STARVATION, AS WELL AS TREATMENT WITH

TUNICAMYCIN.

CC -1 PTM: O-GLYCOSYLATED.

CC -1 SIMILARITY: CONTAINS 3 THIOREDOXIN DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; M86870; AAA19217.1; -.

DR HSSP; P07237; IMK.

DR MEROPS; C17.002; -.

DR InterPro; IPR000886; ER-target.

DR InterPro; IPR00063; ThioRed.

DR Pfam; PF00085; thioRed; 3.

DR PRINTS; PR00421; THIOREDOXIN.

DR TIGRPFAMS; TIGR01126; pdl_dom; 3.

DR TIGRPFAMS; TIGR01130; ER_pdl_fam; 1.

DR PROSITE; PS00014; ER_TARGET; 1.

DR PROSITE; PS00194; THIOREDOXIN; 3.

KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; signal;

KW Glycoprotein; Calcium-binding.

KW Glycoprotein; Calcium-binding.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 643 PROTEIN DISULFIDE ISOMERASE A4.

FT DISULFID 89 92 REDOX-ACTIVE (BY SIMILARITY).

FT DISULFID 204 207 REDOX-ACTIVE (BY SIMILARITY).

FT DISULFID 553 556 REDOX-ACTIVE (BY SIMILARITY).

FT SITE 640 643 PREVENT SECRETION FROM ER.

FT SEQUENCE 643 AA; 72806 MW; 856116663A4DE41 CRC64;

SO

Query Match 3.4%; Score 142; DB 1; Length 643;

Best Local Similarity 22.8%; Pred. No. 0.086;

Matches 115; Conservative 76; Mismatches 201; Indels 112; Gaps 27;

QY 15 VINCIFMPTVNSLPDELSPQKFTSLQGLLENEAVRPLDQGISVAKVNCVKERISKY 74

DB 194 ILVFEYAPWCCHCKALP-----YEKAKELSKRSPP-----IPLAKVDAT----- 236

QY 75 CKEKEDKAVLYFKNNILIREPTDTLPDVN-----AIVAHVLFALLFESEVKYITNED 128

DB 237 --EQDTLAKRPVSGFTPLIKIFRKGRPPYNGPREKYGVDMVWESGPPSKELIL----- 290

QY 129 LQNIENALK-KRANIFSYVAIGIPERHRAVNAEAGFYVGTYYQF--VLTEITA--LLES 183

DB 291 LKQVGEFLKDGDDVILVFGQGVDPGYIQYQDANATLREDYKFIHHTFSTELAKFLKXSL 350

QY 184 GSEDEVYAHLYFPHCKLVL-----DLTQOCRRPLMEOPLTTL--LNHLFITMKAPLI-- 234

DB 351 G-----KLVIAPKEFKQSKYEPKMHVMDVQGSTEASAIKDYVVKHDLPLVG 396

QY 235 -TEVAEDPOQVSTVHLDGLPIYVISOATYREADRTAEVAMRLGKAGVILLRLDSL 293
 Db 397 HRTKTSNDKRYSK-----RPLV-VVYYSVDFSDRTAOF-WR---NKVLEVADFP 444
 QY 294 EVNIPDANVFRKAEVGVVEFLVLDVLIISHVENNHIEIOEDNDMGPDIDV 353
 Db 445 E-----YTFALADE-----EDVATEVKRLGLSEGEEDNMAILDGSKFAMEPE-ER 491
 QY 354 QDDEVAETVFRDRKRKL-----PLELTVELTEETFNATV--ASDSITLV 396
 Db 492 DSDALREEVAFKKGLKLPVKISQPVKNNKGPVRVYV---GKTFDAIVDPKDLIEF 548
 QY 397 YAGWQVSAFLOSIVDVAVKLGSTMLTRINCADMSVCF-KQVATEPFLIKMKKG 455
 Db 549 YAWGCHCKQLEPVYSLGKKRYGQKDLVIAKMD-ATANDITDRYKVGEGPTIFAPSG 607
 QY 456 E--NPVSYAGMLGTKD---LLKFL 474
 Db 608 DKNPIKFEK--GNRDLHLSKFI 629

RESULT 14

PDAA_HUMAN STANDARD: PRT: 645 AA.

AC P13667;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein disulfide isomerase A4 precursor (EC 5.3.4.1) (Protein ERP-72)
 DE (ERP72).
 GN PDIA4 OR ERP72 OR ERP70.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RA MEDLINE=69359272; PubMed=2549034;
 RA Huang S.-H., Tomich J.M., Wu H., Jong A., Holcenberg J.S.;
 RT "Human deoxycyclidine kinase. Sequence of cDNA clones and analysis of
 expression in cell lines with and without enzyme activity.";
 RL J. Biol. Chem. 264:14762-14768(1989).
 RN [2]
 RP REVISIONS.
 RX MEDLINE=91161636; PubMed=2002068;
 RA Huang S.-H., Tomich J.M., Wu H., Jong A., Holcenberg J.S.;
 RL J. Biol. Chem. 266:5353-5353(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Muscle;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 interchain disulfide bonds in proteins to form the native
 structures.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC -1- SIMILARITY: CONTAINS 3 THIOREDOXIN DOMAINS.
 CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A DEOXYCYTIDINE
 KINASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, J05016; AAA58460.1; -
 DR EMBL, BC000425; AAH00425.1; -
 DR EMBL, BC001928; AAH01928.1; -
 DR EMBL, BC006344; AAH06344.1; -
 DR EMBL, BC011754; AAH11754.1; -

DR PIR; A23723; A23723.
 DR HSSP; P07237; IMEK.
 DR MEROPS; C17.002; -.
 DR InterPro; IPR000886; ER_target.
 DR InterPro; IPR000633; ThioRed.
 DR Pfam; PF00085; ThioRed; 3.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR TIGRFAMs; TIGR01126; pd1_dom; 3.
 DR TIGRFAMs; TIGR01130; ER_pdi_fam; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 3.
 KM Redox-active center; isomerase; Endoplasmic reticulum; Repeat; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 645
 FT DOMAIN 39 55
 FT DISULFID 91 94
 FT DISULFID 206 209
 FT DISULFID 555 558
 FT SITE 642 645
 SQ SEQUENCE 645 AA; 72932 MW; 1919C2AE12D2684 CRC64;

Query Match

3.3%; Score 138.5; DB 1; Length 645;
 Best Local Similarity 17.7%; Pred. No. 0.15;

Matches 74; Conservative 72; Mismatches 139; Indels 133; Gaps 13;

QY 327 SHVENNMHIEIOEDNDMGPDIDVODD----- 356
 Db 32 SNENALIEDEEEEDDEEDDELEKENGVLNDANFNVADKDTLLEFYAPWC 91
 QY 357 -----EVAEYFRDRKRKLPL----- 372
 Db 92 GHCKQPAPEYERKIANILKQDPPIVAKIDANSAVLSRFDVSGYPTIKILKGAQVDY 151
 QY 373 -----ELTVELTEETFNATVMSDSITLV-FYAGWQVSA 406
 Db 152 EGSRTQEIYAKVREVSQDPWPPEVTLVLRKENDDEVVNDADILVEFYAPWCHCK 211
 QY 407 FLOSIVDAVKL-KGSTMILTRINCADMSVCF-KQVATEPFLIKMKKGENVASAGML 465
 Db 212 LAPEYERAKELSKRSPPIPLAKVDVTAFTDLAKRPDVGSPYPLKIFRRG-RPYDNGPR 270
 QY 466 GTFDDILQNLNISYPVNTSIOEAEYLSGLRYDLDLYSSVSLGLGFSPTMKAKED 525
 Db 271 EKYGIVDYMEOGSPPSKELITIKOVDFL-----KD---GDDVITIGFKGESPAYQ 322
 QY 526 FSEAGNYLK-GVYITGISEEDVLIISTKYAASLPALLARHTEGIESIPLASTHAODI 584
 Db 323 YQDAANNLREDYKFHHFTFTE-----IAKFLVSGQLVVMQPE-KFQSKYEPESHMDV 376
 QY 585 VQITTDALIEPPEITVENLP-----SYRLOKPLILFSDGTVPQYKKA 630
 Db 377 QGSTQSAIKDF--VLKVALPLVGHKRVGNDAKRYR--RPLVVVYVDFSDFYRAA 430

RESULT 15
 ER60_SCHMA
 ID ER60_SCHMA STANDARD: PRT: 484 AA.
 AC P38658;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Probable protein disulfide isomerase ER-60 precursor (EC 5.3.4.1)
 DE (ERP60).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Liberian;
 RX MEDLINE=94359529; PubMed=8078516;
 RA Finken M., Sobek A., Symmons P., Kunz W.;
 RT "Characterization of the complete protein disulfide isomerase gene of

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:11:22 ; Search time 36 Seconds
(without alignments)
4618.889 Million cell updates/sec

Title: US-09-847-046-2
Perfect score: 4142
Sequence: 1 MFGSFNVFVGVISFVIMCIF.....FKEAKSFRDRKELGCSKVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
1: SPREMBL_21:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.vertibrate:*
15: sp.unclassified:*
16: sp.rvirts:*
17: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4025	97.2	806	4 Q9P2K2	Q9P2K2 homo sapien
2	1839	44.4	357	4 Q9H9W7	Q9H9W7 homo sapien
3	1423	34.4	354	11 Q9CS82	Q9CS82 mus musculu
4	815	19.7	225	11 Q8R2W8	Q8R2W8 mus musculu
5	184.5	4.5	502	3 Q74568	Q74568 trichoderma
6	182	4.4	546	10 Q9C818	Q9C818 arabidopsis
7	178	4.3	519	10 Q9LW75	Q9LW75 arabidopsis
8	176	4.2	489	5 Q76945	Q76945 fasciola he
9	169.5	4.1	482	5 Q26593	Q26593 schistosoma
10	168.5	4.0	505	11 Q9JLF6	Q9JLF6 mus musculu
11	166.5	4.0	505	11 Q91Z81	Q91Z81 cricetus
12	162	3.9	577	16 Q8R6N2	Q8R6N2 thermoaer
13	158	3.8	369	13 Q91815	Q91815 fuigu rubrip
14	154	3.7	494	5 Q9GPH2	Q9GPH2 bombyx mori
15	150.5	3.6	517	3 Q9C128	Q9C128 pichia past
16	150	3.6	579	10 Q8VX13	Q8VX13 arabidopsis

17	148	3.6	566	10 Q9SV44	Q9SV44 arabidopsis
18	147.5	3.6	483	5 Q9GR12	Q9GR12 plasmodium
19	144.5	3.5	359	10 P93358	P93358 nicotiana t
20	143.5	3.5	2471	12 Q9YTK3	Q9YTK3 ateline her
21	141	3.4	416	5 Q01492	Q01492 caenorhabdi
22	138.5	3.3	436	5 Q9VYV3	Q9VYV3 dirosophila
23	138.5	3.3	487	3 Q96VF3	Q96VF3 uscilloago ma
24	137	3.3	366	10 Q93V71	Q93V71 oryza sativ
25	136	3.3	930	16 Q9CUX4	Q9CUX4 pasteurella
26	135	3.3	497	5 Q76191	Q76191 dirosophila
27	134.5	3.2	440	10 Q48773	Q48773 arabidopsis
28	133	3.2	572	5 Q44508	Q44508 caenorhabdi
29	132.5	3.2	2795	2 Q9RNB1	Q9RNB1 microcystis
30	131.5	3.2	574	5 P91442	P91442 caenorhabdi
31	130.5	3.2	363	3 Q13704	Q13704 schizosacch
32	130.5	3.2	485	5 Q17967	Q17967 caenorhabdi
33	130.5	3.2	509	11 Q8R402	Q8R402 cricetus
34	128.5	3.1	2149	10 Q9M3D3	Q9M3D3 arabidopsis
35	128	3.1	977	10 Q9X114	Q9X114 arabidopsis
36	127.5	3.1	480	5 Q96460	Q96460 schistosoma
37	127.5	3.1	488	5 Q17908	Q17908 caenorhabdi
38	127.5	3.1	489	5 Q9TWZ1	Q9TWZ1 dirosophila
39	127.5	3.1	509	11 Q922C8	Q922C8 mus musculu
40	127	3.1	443	10 Q9MAU6	Q9MAU6 arabidopsis
41	126.5	3.1	1014	2 Q8RN21	Q8RN21 campylobact
42	126.5	3.1	1014	2 Q8R1Y7	Q8R1Y7 campylobact
43	126.5	3.1	1214	17 Q8U3H2	Q8U3H2 pyrococcus
44	126	3.0	1042	1 Q9C4Y4	Q9C4Y4 sulfolobus
45	125.5	3.0	371	10 Q942L2	Q942L2 oryza sativ

ALIGNMENTS

RESULT 1									
Q9P2K2		PRELIMINARY;	PRT;	806 AA.					
ID	Q9P2K2								
AC	Q9P2K2								
DT	01-OCT-2000 (TREMBLrel. 15, Created)								
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)								
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)								
DE	KIAA1344 protein (Fragment).								
GN	KIAA1344.								
OS	Homo sapiens (Human).								
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;								
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
OX	NCBI_TaxID:9606;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	TISSUE-BRAIN;								
RX	MEDLINE-20181126; PubMed-10718198;								
RA	Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;								
RT	"Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro."								
RL	DNA Res. 7:65-73(2000).								
DR	EMBL: AB037765; BAA92582.1; -								
DK	InterPro: IPR000063; Thioired.								
FT	NON_TER	1							
SO	SEQUENCE	806 AA;	91407 MM;	B9A4847F928662E3	CRC64;				
Query Match									
Best Local Similarity		97.2%;	Score 4025;	DB 4;	Length 806;				
Matches 788;		Conservative	0;	Mismatches	0;	Indels	18;	Gaps	1;
QY	20 FYMPVNSLPFELSPQKYSTLOPG-----LEELNEAVRPLQDYGISV	61							
DB	1 FYMPVNSLPFELSPQKYSTLOPGKASLAFQADSPRTSVLEELNEAVRPLQDYGISV	60							
QY	62 AKVNCVKEEISRYCKEKDLMKAYLFKGINLREPTDTLFDVNAIVAAVLFALLFSEVK	121							
DB	61 AKVNCVKEEISRYCKEKDLMKAYLFKGINLREPTDTLFDVNAIVAAVLFALLFSEVK	120							

QY 122 YTNLEDLQNIENALCKANIIESYVRAIGIPEHRAVMEAGFYVGTGYQVLTTEIALLE 181
 DB 121 YTNLEDLQNIENALCKANIIESYVRAIGIPEHRAVMEAGFYVGTGYQVLTTEIALLE 180
 QY 182 STSEDEVAHLYEFHCKLVLDLTQOCRRLEMOPLTTLNHLFIKMKAPLITLEVADP 241
 DB 181 STSEDEVAHLYEFHCKLVLDLTQOCRRLEMOPLTTLNHLFIKMKAPLITLEVADP 240
 QY 242 QOVSVAHQGLPLVFTVSQATYEADRTAEVAMRLKAGVLLLRSLSEVNIQDA 301
 DB 241 QOVSVAHQGLPLVFTVSQATYEADRTAEVAMRLKAGVLLLRSLSEVNIQDA 300
 QY 302 NVVFKRAEEVPEVFLVHDVLLIISVENNMHIEEIQDEBDNDMEGPDIDVDEVAET 361
 DB 301 NVVFKRAEEVPEVFLVHDVLLIISVENNMHIEEIQDEBDNDMEGPDIDVDEVAET 360
 QY 362 VPRDRKRLPLELTVLEETFNATVMAISIVFYAGMOVASAFLOSIVDAVVKLGT 421
 DB 361 VPRDRKRLPLELTVLEETFNATVMAISIVFYAGMOVASAFLOSIVDAVVKLGT 420
 QY 422 STMILTRINCADMSVCTKONVTEFPIIKMYKKGENPVSYAGMLGTDLKFTOLNRISY 481
 DB 421 STMILTRINCADMSVCTKONVTEFPIIKMYKKGENPVSYAGMLGTDLKFTOLNRISY 480
 QY 482 PVNITSIQEABEYLSGLYKDLILYSSVYLGFLSPMKTKAKEDFSAGNYLKYVTIGI 541
 DB 481 PVNITSIQEABEYLSGLYKDLILYSSVYLGFLSPMKTKAKEDFSAGNYLKYVTIGI 540
 QY 542 YSEEDVLLSTKYAASLPALLARHTEGKIESIPLASTHODIYOITDALLMEFPIYV 601
 DB 541 YSEEDVLLSTKYAASLPALLARHTEGKIESIPLASTHODIYOITDALLMEFPIYV 600
 QY 602 ENLPSTFRLOKPLILFSDGTNVNPOYKKAITLVKOKYLDSPFCWMLNKNTPVGRILR 661
 DB 601 ENLPSTFRLOKPLILFSDGTNVNPOYKKAITLVKOKYLDSPFCWMLNKNTPVGRILR 660
 QY 662 AYDPLRPLPLVLYNHSGGVAFAPSDAIIIEENLVMLKKEAGLENNHTITLPAQEW 721
 DB 661 AYDPLRPLPLVLYNHSGGVAFAPSDAIIIEENLVMLKKEAGLENNHTITLPAQEW 720
 QY 722 KPLPAYDFLSMIDATSOGRTRKVPKCMKETVOENDKQEHEDKSAVRKEPIETLRIKH 781
 DB 721 KPLPAYDFLSMIDATSOGRTRKVPKCMKETVOENDKQEHEDKSAVRKEPIETLRIKH 780
 QY 782 WNRSMNFKEAESFRDKELGCSKVN 807
 DB 781 WNRSMNFKEAESFRDKELGCSKVN 806

RESULT 2
 ID 09H9W7 PRELIMINARY; PRT; 357 AA.
 AC 09H9W7;
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CDNA FLJ12501 f1s, clone NTMRM200161.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 RA Matsunawa H., Ishii S., Kawai Y., Saio K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK022563; BAB14101.1; -
 SQ SEQUENCE 357 AA; 40469 MW; A7FCD05C2D2E560 CRC64;

Query Match 44.4%; Score 1839; DB 4; Length 357;

Best Local Similarity 99.7%; Pred. No. 2,3e-123;
 Matches 356; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 451 MYKKGENVSYAGMLGTDLKFTOLNRISYPVNITSIQEABEYLSGLYKDLILYSSVS 510
 DB 1 MYKKGENVSYAGMLGTDLKFTOLNRISYPVNITSIQEABEYLSGLYKDLILYSSVS 60
 QY 511 VLGFLSPMTAKEDFSAGNYLKYVITGISEEDVLLSTKYAASLPALLARHTEGK 570
 DB 61 VLGFLSPMTAKEDFSAGNYLKYVITGISEEDVLLSTKYAASLPALLARHTEGK 120
 QY 571 IESTPLASTHODIYOITDALLMEFPIYVENVLPSTFRLOKPLILFSDGTNVPOYKKA 630
 DB 121 IESTPLASTHODIYOITDALLMEFPIYVENVLPSTFRLOKPLILFSDGTNVPOYKKA 180
 QY 631 ILTVKOKYLDSPFCWMLNKNTPVGRILRAYFDPLPLVLYNHSGGVAFAPSD 690
 DB 181 ILTVKOKYLDSPFCWMLNKNTPVGRILRAYFDPLPLVLYNHSGGVAFAPSD 240
 QY 691 QAIIEENLVMLKKEAGLENNHTITLPAQEWKPLPAYDFLSMIDATSOGRTRKVPKM 750
 DB 241 QAIIEENLVMLKKEAGLENNHTITLPAQEWKPLPAYDFLSMIDATSOGRTRKVPKM 300
 QY 751 KETDVOENDKQEHEDKSAVRKEPIETLRIKHNRSMNFKEAESFRDKELGCSKVN 807
 DB 301 KETDVOENDKQEHEDKSAVRKEPIETLRIKHNRSMNFKEAESFRDKELGCSKVN 357

RESULT 3
 ID 09CS82 PRELIMINARY; PRT; 354 AA.
 AC 09CS82;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 5730420B22RIK protein (Fragment).
 GN 5730420B22RIK.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glassi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsch G.,
 RA Blake J., Boffelli D., Bojunga N., Carcinici P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Hayashizaki Y., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Wyshaw-Boris A.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK017582; BMB30819.1; -
 DR MGD; MGI:1917811; 5730420B22RIK.
 FT NON_TER 354
 SQ SEQUENCE 354 AA; 39924 MW; 42B75F03910AEC75 CRC64;

Query Match 34.4%; Score 1423; DB 11; Length 354;
 Best Local Similarity 77.6%; Pred. No. 1,3e-93;
 Matches 274; Conservative 34; Mismatches 27; Indels 18; Gaps 1;

QY 513 GLFSPMTAKEDFEAGNYLKGVIYIT-----GIYSEEDVL--LTSTKYAASLPAL 561
 Db 243 KLFEPNLMT-----NNVFGLVKTETAKTYSYDGPQAKIYEFPLNSKNFP----- 288
 QY 562 LIARHTEG-----KIESIPLASTH-----ADIVQIITDALLEPEITVEN 603
 Db 289 LVTKLTESVAVVSSPVKIQVAFVFSKTDDESLAQPLEDIARKFKSKMLIYIDISNEN 348
 QY 604 LPSYRLQKPLILFSDGTVPQYKKAILLTVKOKYLDSPFCWMLNKTPTPG--RGILR 661
 Db 349 -----LAMPFLTLFGIEDAKTVAAPFNNLNLSKYLESDESPSIFECFGLAHGTVS 402
 QY 662 AVE--DPLPP--LPLVLVNLHSGGVAFPPSDQAI--ENNVLMLKLE---AGLEN 711
 Db 403 AYKKSQPIPDNONASVAVV-----GRTP---DEVVLSSENVLLLEVKLQSHKRGFEN 453
 QY 712 HI--TILPAQEKPPPLPAYDPLSMIDATSQRGTRKVRKCKMETDVQENDKQHEKSA- 768
 Db 454 LVFARIDASANEHPKLTVDYPTIILYKTGE-----KENPLKLTSSAKDMAVL 503
 QY 769 VRKE 772
 Db 504 INKE 507

RESULT 8

ID 076945 PRELIMINARY: PRT: 489 AA.
 AC 076945:
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Protein disulphide isomerase.
 GN PDI GENE.
 OS Fasciola hepatica (liver fluke).
 CC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomidae; Echinostomata; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Salazar M., Casals R., Diaz A., Martin J.M., Boga J.A., Parra F.;
 RT "Cloning and heterologous expression of Fasciola hepatica putative
 RT protein disulphide isomerase";
 RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ225804; CAAL2644.1;
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 3.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 DR Isomerase; Redox-active center.
 SQ SEQUENCE 489 AA; 55222 MW; 047AF15027C872A4 CRC64;

Query Match 4.2%; Score 176; DB 5; Length 489;
 Best Local Similarity 23.2%; Pred. No. 0.00035;
 Matches 65; Conservative 60; Mismatches 99; Indels 56; Gaps 12;

QY 376 VELTEFTNATVMSD-STVLFYAGMOAVSMAFLQSYIDVAVKLKGT-STMLLTRINCAD 433
 Db 31 VELTEFTDEDEKKKEFAVMYAPMGCHCKAMKPEYARAAQOLKEEGSDINIAKVATO 90
 QY 434 WSDVCKQNVTEPPIIKMKYKGENPVSYAGMLGTRDLKFLQJNRIISYVNTS-IOEAE 492
 Db 91 HSKLAKSHNVGTGPTLKFKKSGV-WLDYTGROGKTEIYIMIR-RKVSAPVSLSLSEVQ 148
 QY 493 EYLSEELKDLILYSSVYLGLSPPTMTAKDFSEBAGNYLKGVIYITGIYSEEDVLLST 552
 Db 149 QLVDEK--DIVV-----IAFAEESNEELKOLLE--AVASVYDKYEGFVSS 190

QY 553 KYA-----ASLPALLARHTEGKIESIPLASTHAQDIVQIITDALLEPEITVENLP 605
 Db 191 KQAFDHYKIDSKSRVLEFKKFDGR-----ADPQGLTRELILFQKQETI----- 236
 QY 606 SYFRLQKPLILFSDGTVPQYKKAILLTVKOKYLDSPFCWMLNKTPTPG--RGILR 661
 Db 237 -----PLVETQTOETASAVFGSAI-----RKHYVSVFV 264

RESULT 9

ID 026593 PRELIMINARY: PRT: 482 AA.
 AC 026593:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Protein disulphide isomerase homologue precursor.
 OS Schistosoma mansoni (Blood fluke).
 CC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeiidae;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LIBERIA;
 RX MEDLINE=94359529; PubMed=8078516;
 RA Finken M., Sobek A., Symmons P., Kunz W.;
 RT "Characterization of the complete protein disulphide isomerase gene of
 RT Schistosoma mansoni and identification of the tissues of its
 RT expression";
 RT Mol. Biochem. Parasitol. 64:135-144 (1994).
 DR EMBL: Z22933; CAA80520.1;
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 2.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 DR Isomerase; Redox-active center; Signal.
 FT SIGNAL 1
 SQ SEQUENCE 482 AA; 54154 MW; EA2DE2E2BB209658 CRC64;

Query Match 4.1%; Score 169.5; DB 5; Length 482;
 Best Local Similarity 22.3%; Pred. No. 0.00099;
 Matches 117; Conservative 70; Mismatches 170; Indels 167; Gaps 29;

QY 16 INCIEFMPYVNSLPELSPQKYEFTLOPGLLEINAVRPIQDYG--TSVAKVNC-VKEEIS 72
 Db 43 VLVEFYAPMGCHCKALAP-----EYSEAKKIKLKESGLKAKVQATVEEELA 90
 QY 73 RYCGKAKDLMAKLVKFGNILLRFPPTDL--FDVNAIVAHVLFALLFSE--VKYITNLED 128
 Db 91 LKHG-EKGYPTLKFFR-----NEQPIDFGERSDAIVN--WCLKRSKPSVEYIDLS 141
 QY 129 LQNIENALSKANI-IFSYVR--ATGIRPHAVM-----EAGRVG-----TTYQFVL 174
 Db 142 CKQFTD-----KANIALGLGKTQDSLDLADFEKVADELDDAGFAANSSEILTEYITOT 197
 QY 175 TETALLESIGSEDEVEYAHLYFFHCKLVLDLTQCCRTLMEOPLTTLNLFITKTMAPLL 234
 Db 198 PRTVLFKNPDENKREVT-----GGTLENLKHFTQVESVPLV 233
 QY 235 TEVAEDPQGVSTVHLQGLP-----LVFIVSQATVE--ADRTAEVNAVRLGKAGVLL 288
 Db 234 SEESQ-----KTAGVYEGSPICKHIVFLSKSTSDHSDLDVDKLLE--VARQFGKHLVIV 286
 QY 289 LKDSLEVINIPQANVYFKRAEGSVPEVFLVLDVLDLIIHVENNHTEE----- 337
 Db 287 -----DVD-----VENNLVLEFFGLSKNDAPT 309
 QY 338 ---IQDEEDNDMEGPD-----IDVDDEVAETVFRKRKKLPLELIVE-----L 378

```
DB 310 YRIELGEETTKYKPDNDYSVSAMSDFVORTI--DGKVK-PFLMSEIPSDDTGAVKVL 366
OY 379 TEETFNATV--MASDSIVLFYAGQAVSMAPLOSYIVAVVAKTSTMLLRICADMSD 436
DB 367 VGRKYNVVDKSDVAVKLAAPWCCHCKALAPWDELGEFFKSDVIAK-----MD 419
OY 437 VCTKQ---NYTEFPPIKMY-KKGENPVSYAGMLGTRDLKFLIQ 475
DB 420 ATVNEVEDLKVTSPPLTKFYPKNSEVIDYTGDRSFALKKFVE 463

RESULT 10
O99LF6 PRELIMINARY; PRT; 505 AA.
AC O99LF6;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Similar to ER-60 protease.
GN GRP58.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003285; AAH03285.1; -.
DR HSSP; P07237; IMK.
DR MGD; MG1:95834; GIP58.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 2.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRFSMS; TIGR01130; ER_PDI_fam; 1.
DR TIGRFSMS; TIGR01126; pdl_dom; 2.
DR PROSITE; PS00194; THIOREDOXIN; 2.
KW Protease; Redox-active center.
SQ SQUIDNCE 505 AA; 56678 MW; 3A7CD1C35981C4B3 CRC64;

Query Match 4.1%; Score 168.5; DB 11; Length 505;
Best Local Similarity 22.1%; Pred. No. 0.0013;
Matches 64; Conservative 55; Mismatches 125; Indels 45; Gaps 10;

OY 376 VELTEETFNATVNASDS---IVLFYAGQAVSMAPLOSYIVAVVAKTSTMLLRINC 431
DB 28 LELTDEFESRVSDDTGSAGLMLVEFFAPWCCHCKRLAPEYEAATRLKG--IYPLAKVDC 85
OY 432 ADMSDVCATKNTVEFPPIKMYKKGENPVSYAGMLGTRDLKFLQNLNRPVNTSIQEA 491
DB 86 TANTNTCNKYGVSGLPPLTKFYRDEEAGAYDGPRTADGIVS--HLKKQAPASVILRTEE 143
OY 492 EEVLSGELYKDLILYSSVYLGLFSPMTAKADESEAGNYLKG---YVITGIYSEEDVL 548
DB 144 E-----FKKFSIDKASVVGFRDLFSDGSEFLKAASNLNRYNFAHTNIES----- 191
OY 549 LLSKTYAASLPALLAR--HTEKIESIPLASTHAQ---DIVOITDALLEMPETIVE 602
DB 192 -LVKEYDNGEGITIFRPLHLANKFEDKTVAYTEKKWTSGIKKFIQDSIFGLCPHMTED 250
OY 603 N-----LPSYFRLOKPLILFSDGTVPQYKKAITLVLYKOKYIDS 642
DB 251 NKDLIOGKDLITAYVDV-----YEKNAGKSNYMRNRVMVAKKFLDA 293

RESULT 11
O91281 PRELIMINARY; PRT; 505 AA.
AC O91281;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
```

```
DE ERP57 protein.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue-Ovary;
RC Chung J.Y., Hwang S.O., Lee G.M.;
RT "The nucleotide sequence encoding the chinese hamster ERP57 protein.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057063; AAL18160.1; -.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 2.
DR TIGRFSMS; TIGR01130; ER_PDI_fam; 1.
DR TIGRFSMS; TIGR01126; pdl_dom; 2.
DR PROSITE; PS00194; THIOREDOXIN; UNKNOWN.2.
SQ SEQUENCE 505 AA; 56796 MW; 5F2E38C30794DB76 CRC64;

Query Match 4.0%; Score 166.5; DB 11; Length 505;
Best Local Similarity 21.7%; Pred. No. 0.0017;
Matches 62; Conservative 57; Mismatches 128; Indels 39; Gaps 9;

OY 376 VELTEETFNATVNASDS---IVLFYAGQAVSMAPLOSYIVAVVAKTSTMLLRINC 431
DB 28 LELTDEFESRVSDDTGSAGLMLVEFFAPWCCHCKRLAPEYEAATRLKG--IYPLAKVDC 85
OY 432 ADMSDVCATKNTVEFPPIKMYKKGENPVSYAGMLGTRDLKFLQNLNRPVNTSIQEA 491
DB 86 TANTNTCNKYGVSGLPPLTKFYRDEEAGAYDGPRTADGIVS--HLKKQAPASVILRTEE 143
OY 492 EEVLSGELYKDLILYSSVYLGLFSPMTAKADESEAGNYLKG---YVITGIYSEEDVL 551
DB 144 E-----FKKFSIDKASVVGFRDLFSDGSEFLKAASNLND---NYNFAHTNIESLV 193
OY 552 TKTYAASLPALLAR--HTEKIESIPLASTHAQ---DIVOITDALLEMPETIVEN-- 603
DB 194 KEYDDNGEGITIFRPLHLANKFEDKTVVYEQWTSGLIKRFQESIFGICPHMTEDND 253
OY 604 -----LPSYFRLOKPLILFSDGTVPQYKKAITLVLYKOKYIDS 642
DB 254 LIOSKDLITAYVDV-----YEKNAGKSNYMRNRVMVAKKFLDA 293

RESULT 12
O8R6N2 PRELIMINARY; PRT; 577 AA.
AC O8R6N2;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE ABC-type multidrug/protein/lipid transport system, ATPase
DE component.
GN MDIb12 OR TTE2769.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013214; NAM25873.1; -.
KW Complete proteome.
SQ SEQUENCE 577 AA; 66097 MW; BA6642DF1FBE7545 CRC64;

Query Match 3.9%; Score 162; DB 16; Length 577;
```

Best Local Similarity 19.6%; Pred. No. 0.0045;
Matches 142; Conservative 111; Mismatches 224; Indels 248; Gaps 33;

```

OY 82 MKAYLFKGNILIREPDTLEFVNVAIVAHVLFALLFSEVKYTTNEDLQINIMALKGRKN 141
DB 1 MKEFLFK---YKKEVLAVFLVAVSLINITYAFYKTKLIDVATRTDLKFSYV--KFA 55
OY 142 IIFSIVRAIGLPEHRAVEAGFVYGT--YQFVLTEI---ALLSISEGEVYAHLYF 196
DB 56 VMFTILEAV-----VGLNRTTRHHYKTKLILYKDLFRALIKKDEH----- 99
OY 197 HCKLVLDLTQCCRTLMRQPLTTLNHLFIKTKMAPLLEVAEDPOQV----- 244
DB 100 -----FNEVNTGKIYISINDYKIVVEDYFNNFFRL 131
OY 245 -STVHLQGLPLVFISSQATVEADRRTAEMVAMRLLAGVLLRLDSELENNIQ--DA 301
DB 132 GSANGFVAALLSFLSTKIT-----YMLIMALLSVIIPRIED 171
OY 302 NVFERAEAGVPEFLVLDVLLIS-----HYENNHIIEIOEDENDMEGPDIDVQ 354
DB 172 KIAKRKNYSESELFTIESKDTLGLGVKSFGEIDAKHEKFSKVED----- 220
OY 335 DDEVAETVFRDRKKLPLELVLTETFPNATVMSDSIVLFYAGQAVSMAFLQSYIDV 414
DB 221 -----VEDKKIKYSVLLN---TSDTMSSELISFIFLSVF-----AVGLYF----- 257
OY 415 AVKLKGTSMILTRINCADMSDVCTKQNTPEPIIKMYKKGKNPVSYAGMLGCKDLKRTI 474
DB 258 --TIKGMT-LGTMIACV-----QLTNNIIMPIYSMGON-----LNRTL 293
OY 475 QLNRIYSVNTISIOEAE---YLSGELYKDLILSVSVGLFSPMTKTAKED-----F 526
DB 294 SLKTSQKIN-EVLDKKEKENDIYVKSNDSEIKRNS---FSTGTAKLDNINFTI 348
OY 527 SEAGNYLKVYITGIVSEEDVLLSTKYA--ASLPALLARHT---ECKIESIPLASTH 580
DB 349 KKGKRY-----ALVGTSGAGKSTILKLLKQYENYEGETIKLDGIELRRID 393
OY 581 AODIVQITDALLEMFPEITVENLPSYFRLQPLILPSDGVNPOYKKAITLYKQKLT 640
DB 394 KKDLEKITT-----LLHQN-----VFIFDGTV-----KDNITLFDNRYT 427
OY 641 DSFTPCMLNKTPYRGILRAYFDP-LPPLPLVLVNLHSGGYAFPSDQ-----AI 693
DB 428 DE-----EYVRAKTIAGLGPLLEKLPREGILSDVGECKLISGGERRIARSI 476
OY 694 IEEHLVLMKLKLEAGLENN-----IT--ILPAQEMKPPPLPAYDFLSMT-DAA 737
DB 477 ITNASTIALDEKTAALDMETAYMIKTIIDMTAIVVTHRLMSSELKRYDELIYLRQGR 536
OY 738 TSORG 742
DB 537 IVEKG 541

```

RESULT 13

```

OY91815 PRELIMINARY: PRT: 369 AA.
AC Q91815:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE Protein disulfide isomerase ER-60 (EC 5.3.4.1) (Fragment).
GN ERp60.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OC NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Fu R., Jiang M., Brenner S.;
RT "Structure of the ubiquitous mitochondrial creatine kinase gene and
its linkage to the protein disulfide isomerase Erp60 gene are
RT conserved among teleost fish Takifugu rubripes, mouse and human."
RT Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

```

DR EMBL; AF274502; AAF78087.1; -.
DR HSSP; P07237; 1MER.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR PROSITE; PS00194; THIOREDOXIN; 1.
KW Isomerase; Redox-active center.
FT NON_TER 116 116
FT VARIANT 116 116 T -> S.
SQ SEQUENCE 369 AA; 41237 MW; 1438527CCAD01E55 CRC64;

```

Query Match 3.8%; Score 158; DB 13; Length 369;
Best Local Similarity 22.0%; Pred. No. 0.0043;
Matches 61; Conservative 60; Mismatches 130; Indels 26; Gaps 9;

```

OY 376 VELTEETFNATVMSDSIVL-FYAGQAVSMAFLQSYIDVAVKLKGTSMILTRINCAD 434
DB 21 LEFTDNDFESKIGDHEITLVEFFAPWCGRKRLAPYEKAAATLAKG--VYPLAKVCTSN 78
OY 435 SDVCTKQNTPEPIIKMYKKGKNPVSYAGMLGCKDLKFTQLNRIYSVNTISIOEAEY 494
DB 79 SNICKSYQVSGYPTLKFVFDGESGAYDGPRTSDGIVTFKKQVGPASVALAGEELQKF 138
OY 495 LSGELYKDLILSVSVGLFSPMTKTAKEDPSEAGNYLKG-YVITGIYSEEDVLLSTK 553
DB 139 ISEK-----DSSVGVFFADKSTAQVEFLKASALRDXYRFA--HTNVEVLLKSON 187
OY 554 YVASLPALLARHTCKEISIPLASTH---AODIVQITDALLEMFPEITVENLPSYR 609
DB 188 VQPEGIVLFRPPTLKKKFFEDSSVKYISEKTYTSKIKRFTQDNLGRCPHATEDNKQL-- 245
OY 610 LQKPLILFSDGTV--NPQ---YKKAITLYKQKYL 641
DB 246 MGKDLLVAAYDVYERNRPGSNWNRNWKVAKTFED 282

```

RESULT 14

```

OY9GPH2 PRELIMINARY: PRT: 494 AA.
AC Q9GPH2:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Protein disulfide isomerase.
GN PDI.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Bombycidae; Bombyx.
OC NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RA Goo T.W., Yun E.Y., Hwang J.S., Kang S.W., Lee K.S., Lee J.S.,
RA Kwon O.-Y.;
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF325211; AAG45936.1; -.
DR HSSP; P07237; 1MER.
DR InterPro; IPR000866; ER_target.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 2.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRfams; TIGR01130; ER_PDI_fam; 1.
DR TIGRfams; TIGR01126; pdi_dom; 2.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
DR PROSITE; PS00194; THIOREDOXIN; 2.
KW Isomerase.
SQ SEQUENCE 494 AA; 55588 MW; 084609DD56DC88DD CRC64;

```

Query Match 3.78; Score 154; DB 5; Length 494;
 Best Local Similarity 22.08; Pred. No. 0.013;
 Matches 64; Conservative 54; Mismatches 125; Indels 48; Gaps 11;

QY 369 KLPELTV-ELTEETFNATVWASDIYL-FYAGQAVSMALQSYIDVAVL-KGTSTM 425
 Db 19 EVPEENVLVLSKNETFVISTEYILVERFAPWCGHCKSLAPYAKAAKRLAEESPIK 78
 QY 426 LTRINCADWSDVCTKONVTEPPIIKMYKGENPVSYAGMLGTCKDLKFIQINRISYVNI 485
 Db 79 LAKVDATQEODLAESYGVGRPTLKFFRNG-SPIDYSGRQADDIIMLKKTGPRAVEY 137
 QY 486 TSIOAEETLSELGELYKDLILSSVSLGSPYTKTAKEDPSEGNLKGVIITGISEE 545
 Db 138 TSAEQA-----KELIDANTVIVFGFSDQSTRAKFTLSAOVVDQVF-AIVSDE 187
 QY 546 DVLSTKRYAASLPALLARHTGKIESIPLASTHADIYOITDALLEMPETVLELP 605
 Db 188 KVI-----KELEADEDVVLFKNFEKRYKYEDE-----ETTEDLN 224

QY 606 SYERLQK-PLLIFFSDGTNP-----QYKAILTLVK---QRYLDSFTP 645
 Db 225 AWVVGQSMPTVERSHETASKIFGKIKYHLLIFLSKNGDFEXYLDLKP 275

RESULT 15

09C128 ID 09C128 PRELIMINARY; PRT; 517 AA.
 AC 09C128;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Protein disulphide isomerase.
 GN PDI.
 OS Pichia pastoris (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4922;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21139750; PubMed=11243858;
 RA Wiersma A.A., Vad R., Kristensen T., Oyen T.B.;
 RT "Characterization of a gene encoding a Pichia pastoris protein
 RT disulfide isomerase."
 RL Biochem. Biophys. Res. Commun. 281:1176-1182(2001).
 DR EMBL: AJ302014; CAC33587.1; -.
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000886; ER target.
 DR InterPro: IPR000063; Thioled.
 DR Pfam: PF000085; thioled; 2.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 KW isomerase; Redox-active center.
 SQ SEQUENCE 517 AA: 57788 MW: 40085C534B65583 CRC64;

Query Match 3.6%; Score 150.5; DB 3; Length 517;

Best Local Similarity 22.5%; Pred. No. 0.025;
 Matches 79; Conservative 64; Mismatches 161; Indels 47; Gaps 15;

QY 376 VELTEETFNATVWASDIYL-FYAGQAVSMALQSYIDVAVL-KGTSTM 434
 Db 36 VKLTEATFESFITSNPHVLAFFAPWCGHCKILGPELSAAILKQNEQVIAQIDCTEE 95
 QY 435 SDVCTKONVTEPPIIKMYK-KGENPVSYAGMLGTCKDLKFIQINRISYVNIITGISEE 493
 Db 96 KELCOGEIIGYPTLKFHGEVEPSPDYQGOOSOSIVSYMLKQSLPVPVSEINATKDLD 155
 QY 494 YLSELGYKDLILSSVSLGSPYTKTAKEDPSEGNLKGVIITGISEE--DVLIS 551
 Db 156 TIA-EAKEPVI---VOVL-----PEDASNLESNTTFYGVAGTLREKFTFVSTKS 200

QY 552 TKYAA-----SLPALLARHTGKIESI-----PLASTHADIYOITDALLEMPETIVE 602
 Db 201 TDYAKKRTSDSTPAYLLVR--PGEEPSVYSGEELDETHLVHMIDIESKPL---FGDIDGS 255
 QY 603 NLPYFRLQKPLLIFFSDGTVPNPOYKKAILTLYK-----QRYLDSFPCMLNKNPVG 657
 Db 256 TFKSYAENIPLAVYFE---NEQRAAADIIPKPAKEQRGKINF---VGLDAVKEGK 308
 QY 658 GILRAYDPPLPPLVILVNLHSGQGVFAFPDQALIEENLVLMKLEAG 708
 Db 309 HAKNLNME-EKLPFYIHDLVS-NKKEGVPODELTKNVTETIEKFTAG 357

Search completed: March 13, 2003, 17:14:05
 Job time : 41 secs

```

DB 2401 AATATCTACTGATATCTGTTCTTAATTTCTTCACATATGTTGCTGATACAAAC 2460
OY 4146 TGAATTTTAACTGAATTAAGAAATCTAAACGCTAAACGATGATGATATTT 4205
DB 2461 TGAATTTTAACTGAATTAAGAAATCTAAACGCTAAACGATGATGATATTT 2520
OY 4206 TCCATTAATCATAGACCCGCTGCTACTGACGACCCCTCCCTCAATTTTTTCCGTA 4265
DB 2521 TCCATTAATCATAGACCCGCTGCTACTGACGACCCCTCCCTCAATTTTTTCCGTA 2580
OY 4266 GCATGATGATCCGCTGATTAACGATTTTCATTTTCTTAATATGGAACAAATG 4325
DB 2581 GCATGATGATCCGCTGATTAACGATTTTCATTTTCTTAATATGGAACAAATG 2640
OY 4326 AGATGATGATCCGCTGATTAATGCTGATTAATTAACATCATGATTAATTAAGAAA 4385
DB 2641 AGATGATGATCCGCTGATTAATGCTGATTAATTAACATCATGATTAATTAAGAAA 2700
OY 4386 TGGTAATTAAGTACGACGACATAGAAACATGAATGCTGATGATGATGATGATG 4445
DB 2701 TGGTAATTAAGTACGACGACATAGAAACATGAATGCTGATGATGATGATGATG 2760
OY 4446 CAATTTTGACGATCATTAATGTTGTCATTAATTTTAATTAAGTGTGCTGATGAT 4505
DB 2761 CAATTTTGACGATCATTAATGTTGTCATTAATTTTAATTAAGTGTGCTGATGAT 2820

RESULT 3
LOCUS AK097748 2747 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens CDNA FLJ240429 fls, clone TEST12039177.
ACCESSION AK097748
VERSION AK097748.1 GI:21757613
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens testis CDNA to mRNA, clone lib-TEST12
Clone:TEST12039177.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1
Oshima,A., Takahashi,Fuji,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Matsushima,M., Murekawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
Suzuki,Y., Sugano,S., Negahori,K., Masuno,Y., Negai,K. and
Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2747)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazumi-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan. cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5- & 3-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1. 2747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TEST12039177"
/tissue="testis"
/clone_lib="TEST12"
/note="cloning vector: pME18SFL3"

```

```

BASE COUNT      899 a      510 c      575 g      763 t
ORIGIN
Query Match      53.3%; Score 2410.2; DB 9; Length 2747;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 2598; Conservative 0; Mismatches 3; Indels 156; Gaps 2;
OY 139 GTTGGCGCGCGCGCTCGGGAACGCGCCAGTCCCGCCGCGCGGAGCTTAC 198
DB 1 GTTGGCGCGCGCGCTCGGGAACGCGCCAGTCCCGCCGCGCGGAGCTTAC 60
OY 199 ATGATCATCAGTAGAAGAACTTCTTGAAGTGTTCAGAAATTTGAAGTAGCAAAAT 258
DB 61 ATGATCATCAGTAGAAGAACTTCTTGAAGTGTTCAGAAATTTGAAGTAGCAAAAT 120
OY 259 AGAAATTAAGAAATTAAGCAAGATACAGAGGACACATGGAAGTGTCTTAGAAG 318
DB 121 AGAAATTAAGAAATTAAGCAAGATACAGAGGACACATGGAAGTGTCTTAGAAG 180
OY 319 AGAAGACAGAGAGTGAAGAAACAGACAAATCCGCTCAGATACAGTGCAGTGAATGT 378
DB 181 AGAAGACAGAGAGTGAAGAAACAGACAAATCCGCTCAGATACAGTGCAGTGAATGT 240
OY 379 TTTCCGCTTCAATGCTTTAGAGTTGGATCTTTTGTCAATATGTCATTTTACA 438
DB 241 TTTCCGCTTCAATGCTTTAGAGTTGGATCTTTTGTCAATATGTCATTTTACA 300
OY 439 TGGCAACAGTAACTCTTTACGAGACTGATGCTCTGAGAAATTTTACTACATTCGCAAC 498
DB 301 TGGCAACAGTAACTCTTTACGAGACTGATGCTCTGAGAAATTTTACTACATTCGCAAC 360
OY 499 CAGG-----TC 504
DB 361 CAGGAAAGACCTCTTTAGCTATTTTGTCAAGCTGATTCGCCAGAACATCTGATTTTC 420
OY 505 TTGAAGAACTGAATGAGCTGTTAGACCTCTGACAGCATGTAATTTAGTCCCAAG 564
DB 421 TTGAAGAACTGAATGAGCTGTTAGACCTCTGACAGCATGTAATTTAGTCCCAAG 480
OY 565 TTATTTGTGCAAGAAAGAAATATCAAGTACTGTGGAAGAAAGATTGATGAAG 624
DB 481 TTATTTGTGCAAGAAAGAAATATCAAGTACTGTGGAAGAAAGATTGATGAAG 540
OY 625 CATATTTATTCAGGGCAACATATTCCTGAGAAATTCCTGACACCTGTTTGATG 684
DB 541 CATATTTATTCAGGGCAACATATTCCTGAGAAATTCCTGACACCTGTTTGATG 600
OY 685 TGAATGCCATTTGTCGCCCATGTTCTTCTTTTGTAGTGAAGTGAATATTA 744
DB 601 TGAATGCCATTTGTCGCCCATGTTCTTCTTTTGTAGTGAAGTGAATATTA 660
OY 745 CCACCTGAGAGACCTTGAAGCATGAAGAAATGCTCTGAAGAAAGCAATATTTAT 804
DB 661 CCACCTGAGAGACCTTGAAGCATGAAGAAATGCTCTGAAGAAAGCAATATTTAT 720
OY 805 TCTCATATTAAGAGGACCATTTGAATACGAGACAGACAGTCAATGAAGCCGTTTGG 864
DB 721 TCTCATATTAAGAGGACCATTTGAATACGAGACAGACAGTCAATGAAGCCGTTTGG 780
OY 865 TGTATGGAGCTACATACCAATTTGCTTACACAGAAATTCCTTTTGAAGATATG 924
DB 781 TGTATGGAGCTACATACCAATTTGCTTACACAGAAATTCCTTTTGAAGATATG 840
OY 925 GCTGAGAGATGTGAATATGCAATCTCTTACTTTTCTTATGTAAGTACTGAGT 984
DB 841 GCTGAGAGATGTGAATATGCAATCTCTTACTTTTCTTATGTAAGTACTGAGT 900
OY 985 TGACCCAGCAATGTGAAGACAGTGAAGACAGCAGTCTGACTGACATTCAC 1044
DB 901 TGACCCAGCAATGTGAAGACAGTGAAGACAGCAGTCTGACTGACATTCAC 960
OY 1045 TGTATTAAGCAATGAAGACAGCAGTCTGAGTGAAGTCTGAAATCTCTCAAG 1104
DB 1104 TGTATTAAGCAATGAAGACAGCAGTCTGAGTGAAGTCTGAAATCTCTCAAG 1044

```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

QY 2684 AAAGAACCCGATTGAACCTCTGAGAAATAGAGCATTTGGAATAGAGTAATTTGGTTTAAAGA 2743
 Db 1038 AAAGAACCCGATTGAACCTCTGAGAAATAGAGCATTTGGAATAGAGTAATTTGGTTTAAAGA 1097
 QY 2744 ACAGAAAAATCATTTAGACGCTGATAAAGAGTTAGAGTGTCTCAAAAGTGAACCTAAATTTTA 2803
 Db 1098 ACAGAAAAATCATTTAGACGCTGATAAAGAGTTAGAGTGTCTCAAAAGTGAACCTAAATTTTA 1157
 QY 2804 TAGGCTGTGGTTTCCAAAATTTTGGCATGATAGACATTAATTTATTTTCCCTTAAAGAA 2863
 Db 1158 TAGGCTGTGGTTTCCAAAATTTTGGCATGATAGACATTAATTTATTTTCCCTTAAAGAA 1217
 QY 2864 TAATATTAATCATTTTCAAGCTTGCAGACTAGTGCATCAATAGAGTAATTAATATAAGT 2923
 Db 1218 TAATATTAATCATTTTCAAGCTTGCAGACTAGTGCATCAATAGAGTAATTAATATAAGT 1277
 QY 2924 CACATATTTTATTAATAATTTCTAGTAACATACATTAACCAAAAGTAAAGTGAGCAGGCG 2983
 Db 1278 CACATATTTTATTAATAATTTCTAGTAACATACATTAACCAAAAGTAAAGTGAGCAGGCG 1337
 QY 2984 AAATAATTTTGTATATTTTACCCAGTACTATACCCAAATAGCGAAATATAGAAA 3043
 Db 1338 AAATAATTTTGTATATTTTACCCAGTACTATACCCAAATAGCGAAATATAGAAA 1397
 QY 3044 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 3103
 Db 1398 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 1457
 QY 3104 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 3163
 Db 1458 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 1517
 QY 3164 TGGCTGTGACTACTGATTTGGACAGTTCAGTACTAGACAAAACAGTAACTAACTT 3223
 Db 1518 TGGCTGTGACTACTGATTTGGACAGTTCAGTACTAGACAAAACAGTAACTAACTT 1577
 QY 3224 AGTTCTAGCCATGATTTCTATTTGGATTTAAATTTAACTCTAATCAGAGTTAACTCACA 3283
 Db 1578 AGTTCTAGCCATGATTTCTATTTGGATTTAAATTTAACTCTAATCAGAGTTAACTCACA 1637
 QY 3284 GTGCAATTCATGCACTGACAGTTATATTTGTTTATTTGGAGTCATGATATTAATCAGC 3343
 Db 1638 GTGCAATTCATGCACTGACAGTTATATTTGTTTATTTGGAGTCATGATATTAATCAGC 1697
 QY 3344 GTTGTCAACTCAGGCGATATTTAGCAATTTGCGGAGACATTTTGTATGTCATGACTA 3403
 Db 1698 GTTGTCAACTCAGGCGATATTTAGCAATTTGCGGAGACATTTTGTATGTCATGACTA 1757
 QY 3404 GGCAGTTATTGACATTTAGTAGTAGAGCCATGCTGCTGCTAAATAACCTGCAATTTG 3463
 Db 1758 GGCAGTTATTGACATTTAGTAGTAGAGCCATGCTGCTGCTAAATAACCTGCAATTTG 1817
 QY 3464 ACAGCGCCCAACAAGAAATATCTGCGGAAATGGTAGTGTGCGCAAGGCTGAGTA 3523
 Db 1818 ACAGCGCCCAACAAGAAATATCTGCGGAAATGGTAGTGTGCGCAAGGCTGAGTA 1877
 QY 3524 ACCTTGTGTTAAAGTAGTAACCTGTGGCAGACTAGTGTTCAGAGATTTTCCGTTCTGCTCA 3583
 Db 1878 ACCTTGTGTTAAAGTAGTAACCTGTGGCAGACTAGTGTTCAGAGATTTTCCGTTCTGCTCA 1937
 QY 3584 CGTATCATGTTTGAATAATTTTGGCTATTAAAGATATGATTAGATGGTCTTATCTCGA 3643
 Db 1938 CGTATCATGTTTGAATAATTTTGGCTATTAAAGATATGATTAGATGGTCTTATCTCGA 1997
 QY 3644 TTATTACCTGGATACACTTGAATCTTTTCTAAATTTTTCAGAAAGTATGGGATAACCCCT 3703
 Db 1998 TTATTACCTGGATACACTTGAATCTTTTCTAAATTTTTCAGAAAGTATGGGATAACCCCT 2057
 QY 3704 AGAAGAGGACTCAGATGATATTTTAAAGTGAAGTCTTAAACCTCCTCTTATTTTC 3763
 Db 2058 AGAAGAGGACTCAGATGATATTTTAAAGTGAAGTCTTAAACCTCCTCTTATTTTC 2117

QY 3764 TACAAGTTATATGGCTAAATTTTCAGATTTGAACAGGATTCAGCATTTCTGCCATCTCTCA 3823
 Db 2118 TACAAGTTATATGGCTAAATTTTCAGATTTGAACAGGATTCAGCATTTCTGCCATCTCTCA 2177
 QY 3824 TGAAGAAGAGAGGCTCCCTCATCTGAAGCGTCTCTGAAATCTACCCCTTGCAGCTTCAGAC 3883
 Db 2178 TGAAGAAGAGAGGCTCCCTCATCTGAAGCGTCTCTGAAATCTACCCCTTGCAGCTTCAGAC 2237
 QY 3884 AAATCAGTTGATCTCCCTGAGCCACAGCGCTCATCTGTGAGGGAGGAAAGATAGCC 3943
 Db 2238 AAATCAGTTGATCTCCCTGAGCCACAGCGCTCATCTGTGAGGGAGGAAAGATAGCC 2297
 QY 3944 AAAGAGTTAAATTTTCAATTCACAAATCACTTACCTTACCTGATCTGTTTGTAGCAGTTG 4003
 Db 2298 AAAGAGTTAAATTTTCAATTCACAAATCACTTACCTTACCTGATCTGTTTGTAGCAGTTG 2357
 QY 4004 TTTGTCTCATTTTGTCTGTCATTTTGTGAGACATTTTGTGAGACATTTTGTGAGACATTTTGTG 4063
 Db 2358 TTTGTCTCATTTTGTCTGTCATTTTGTGAGACATTTTGTGAGACATTTTGTGAGACATTTTGTG 2417
 QY 4064 GCTCTACTGATTTTCTTTTAAATATCTACTGATATCTTGTCTTTTAAATTTTCTTCA 4123
 Db 2418 GCTCTACTGATTTTCTTTTAAATATCTACTGATATCTTGTCTTTTAAATTTTCTTCA 2477
 QY 4124 CATATGTTTGCCTGATACAACTGATTTTATTAAGTAAATTTAAGGAATCTAACAGCTA 4183
 Db 2478 CATATGTTTGCCTGATACAACTGATTTTATTAAGTAAATTTAAGGAATCTAACAGCTA 2536
 QY 4184 AAATCAGTAAGTGCAATTTTCTTATTAACATAGACCCCTTGTCTACTCTCAGCACCCCT 4243
 Db 2537 AAATCAGTAAGTGCAATTTTCTTATTAACATAGACCCCTTGTCTACTCTCAGCACCCCT 2592
 QY 4244 CTCTCAATTTTCTCTGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 4303
 Db 2593 CTCTCAATTTTCTCTGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 2647
 QY 4304 TATTCTAAATATGGGAACAATGAGAGTGAACCTCTAAATATAGTTTGTAGTAAATAACAT 4363
 Db 2648 TATTCTAAATATGGGAACAATGAGAGTGAACCTCTAAATATAGTTTGTAGTAAATAACAT 2707

RESULT 4
 ABV27799
 ID ABV27799 standard; cDNA; 2905 BP.
 XX
 AC ABV27799;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 27790.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX

THIS PAGE BLANK (USPTO)
